UNITED NATIONS
ECONOMIC AND SOCIAL COUNCIL

Distr.
LIMITED
ST/ECA/SM/6
15 September 1981
Original: ENGLISH

ECONOMIC COMMISSION FOR AFRICA
Seminar on Household Surveys
Addis Ababa, 21-30 September 1981

LC/NAIROBI

TECHNICAL PLANS FOR SOCIAL SURVEYS

C O N T E N TSS

	Paragraph	Page
Introduction.....	1-4	1-2
Types of surveys.....	5-6	3-4
Concepts, definitions and classifications.....	7-12	4-5
Special problems.....	13-16	5-6
Conclusion.....	17-19	7-8

ANNEX I: Health
ANNEX II: Education and Literacy

Introduction

1. It may be recalled that the African Household Survey Capability Programme (AHSCP) aims at assisting African countries to collect integrated social, demographic and economic data on households and household members. In the proposals so far formulated by ECA, this objective is to be achieved by assisting these countries to establish statistical infrastructures and to formulate and implement integrated programmes of household surveys. Integration is to be achieved through a core questionnaire which is to be administered during each round with specific modules to be added when needed. The fields of priority identified in the region and the data requirements is given in the paper "Data Requirements" E/CN.14/SM/22. These are as slightly modified in the AHSCP Prospectus:

(a) household consumption, expenditure, income, assets, debts;

(b) labour force, employment, unemployment and under-employment

(c) demographic characteristics, household and family structure, position of women and children;

(d) fertility, mortality and migration;

(e) conditions of housing, water supply, health, nutrition, education, literacy, culture, and access to and use of related services;

(f) productive activities and outputs of households including household enterprises (agriculture, industry, trade, etc.) and access to and use of related facilities.

2. It would be noted that information on social conditions can be obtained in respect of all the above subject fields, with the major area being (e).

3. It is now widely recognized that the study of social conditions is important not only for understanding the process of development but also for monitoring changes in the structure of society. In the past emphasis had been placed on the collection of economic data. later this was shifted to demographic data. The need for social data in the context of integrated economic, demographic and social statistics is now generally acknowledged.

4. However the concept of a social survey is more difficult to define. A simple approach would be to consider all surveys concerned with the collection of data on the conditions of and relations among human beings as constituting social surveys. Social surveys can therefore be defined to cover all aspects of social life particularly those of concern at particular points in time. In general there are three approaches to social surveys. An integrated approach or modules on special areas of concern or a combination of both. In order to understand more clearly the vast field covered by social surveys it is necessary to look at some suggestions on the scope. In the United Nations

publication 'Towards a System of Social and Demographic Statistics', 1/ a classification of the subject fields is given as follows:

The size and structure of the population, births, deaths and migration

Family formation, families and households

Social class, stratification and mobility

The distribution of income, consumption, accumulation and net worth

Housing and its environment

The allocation of time and the use of leisure

Social security and welfare services

Learning activities and educational services

Earning activities, employment services and the inactive

Health and health services

Public order and safety, offenders and their victims

This classification contrasts with that given by the US Bureau of the Census 2/ which is as follows:

Population and the family

Health and Nutrition

Housing and the environment

Transportation

Public safety

Education and training

Work

Social Security and welfare

Income and productivity

Social participation

Culture, leisure, and use of time

In Africa the last known example of delimiting subject fields within this vast area of social concerns is the one given in the Kenya document

---

1/ United Nations. 'Towards a System of Social and Demographic Statistics.' United Nations. New York, 1975.

2/ US Bureau of the Census: Social Indicators III. Bureau of the Census. Washington D.C. 1980.

"Towards a Social Perspective: A Statistical Appraisal." 3/ In this document six main areas of concern are discussed and indicators provided. These areas are:

Health and nutrition

Education

Housing

Levels of living

Public safety

Community

Types of surveys

5. In general therefore social surveys in Africa and elsewhere have followed either an integrated approach for which data are collected across the subject fields or a special module approach in which data are collected in relation to a specific module. The United Kingdom Government Social Survey which has now been absorbed by the Office of Population Censuses and Surveys carried out a series of ad hoc surveys commissioned by different departments in their areas of concern such as housing survey, social welfare for the elderly survey on labour mobility and sickness survey. In addition it had a continuing household survey programme from which integrated data could be obtained. In Africa three types of surveys in this field can be given as illustrations. The social statistics and social monitoring survey which was carried out by UNICEF in the Southern Sudan had as its objective the collection of data for monitoring social progress in that part of the country. There was thus emphasis on community level variables (particulars of chief's area) such as name of village, area, language, major crops, corpping seasons and number of permanent wells. It should be mentioned however that this survey does not use the household as the unit of inquiry although information on the number of households in the area is collected. The next example comes from the Kenya National Integrated Sample Survey Programme of 1974-78 where a wide variety of information on the social conditions in rural Kenya was collected. In particular the modules that were run included the following:

Household amenities

Shopping habits

Nutrition

Rural non-farm activities

Handicapped population

Non-formal education

The division of labour and

Health module

3/ Kenya Central Bureau of Statistics: Towards a Social Perspective: A Statistical Appraisal. Central Bureau of Statistics, Nairobi, Kenya.

It was possible from the main survey itself and from the modules to get comprehensive information on social conditions which could be used not only to measure the state of social progress but also to monitor this. The third example is that of Botswana's informal sector survey which covered three peri-urban areas of Botswana (Gaborne, Selebi-Pikwei and Francistown) in 1974. The survey collected information for the study of population size, size of households, age distribution, sex distribution, migration and village of origin, year of arrival, reasons for coming to and staying in the peri-urban areas, education and training, work status, unemployment, gainful employment, place and type of work, type of housing, visits to health care facilities, child mortality fertility and household incomes. The purpose of the survey was to collect data which were urgently required for the formulation of policies on peri-urban development.

6. It would appear that while the first example deals generally with the whole field of social conditions the last two examples follow more or less the United Kingdom example in concentrating on special areas of concern.

Concepts, definitions and classifications

7. In Annexes 1 and 2, detailed plans for health, and education and literacy surveys are discussed. Of general concern is the concepts, definitions and classifications relating to education, literacy, educational attainment, school attendance and disability. It is not here intended to repeat the discussions in the Annexes. This section will focus mainly on the general problems usually encountered in social surveys.

8. While questions of concepts, definitions and classifications in the economic and demographic field are difficult they appear to be even more intractable in the field of social surveys.

9. In the context of household surveys, the measurement of disability is not easy. In the case of blindness which is not total, it is not easy to reflect the degree of blindness through the interview method unless this is accompained by careful screening by specially trained personnel. When such information is supplied by a person, who is not the respondent, the degree of reliability decreases appreciably. In the case of mental illness, most African households will not divulge information on household members who are mentally sick.

10. It may hawever be mentioned that Madagascar and Ghana have tried to collect information on the disabled during their population censuses. Madagascar asked a straightforward question on disability. Ghana's approach was to collect information on those aged 10 years or more who were unable to work because of disability. Both approaches appear to have underestimated the true size of the disabled population.

11. In the household amenities module used in connexion with the Kenyan Rural Survey (1976-77), questions were asked with respect to source, distance and fetcher of water. In respect of the second information,

the classification used was (in kilometers): 0, 1, 1-2, 2-4, 4-8 and 8 . It is not clear how the respondent was able to indicate the distance and how accurate were the recorded responses. Experience from other countries would however suggest that unless careful instructions are given to interviewers, the collected information is not very useful.

12. With respect to housing, some information has been collected over the years during population censuses. It is thus possible to consider some of the conceptual problems based on extensive experience in the region. These problems are discussed in "Study on Methods and Problems of the 1970 Round of African Population and Housing Censuses" E/CN.14/CAS.10/15.4/ The items discussed include number of houses, type of housing unit, tenure, material of wall, material of roof, material of floor, source of water supply, light and fuel, type of toilet, bathing facilities and number of rooms.

Special problems

13. A number of problems are encountered, some of which are either special to or exaggerated in social surveys. The first is poor response in this type of survey. Non-response is not generally a serious problem in rural areas where appropriate contacts with village headmen is usually enough to ensure high response rates. In urban areas, however, there is some difficulty. This is not a peculiar problem in developing countries. It is even more pronounced in developed countries. Some attempts have been made to overcome this in social surveys by using monetary incentives to reduce non-response substantially and at the same time increase cost effectiveness. One such attempt is discussed by Chromy and Horvitz.5/ A national household survey was conducted in the United States to study the level of knowledge of persons aged 26-35 years in the areas of citizenship, science and writing. In this survey special exercises were to be administered to eligible respondents. The response rate in the first year was very poor - 44.4 per cent. In the second year, a controlled experiment using four procedures were tried. These were

a) No monetary incentive

b) Five-dollar incentive to take a package (i.e., one of ten sets of examination questions)

c) Ten-dollar incentive to take a package

d) A variable incentive package.

4/ UN Economic Commission for Africa, Study on Methods and Problems of the 1970 Round of African Population and Housing Censuses. Paper presented to the tenth session of the Conference of African Statisticians, Addis Ababa 1977.

5/ Chromy, James R. and Horvitz, Daniel G. (1978). The Use of Monetary Incentives in National Assessment Household Surveys. Journal of the American Statistical Association 73, 473-478.

It was shown that the variable incentive approach made it possible for respondents to provide more information without complaining about the time required - one hour for each package. In rare populations, the use of incentives could actually reduce the cost per unit of the information collected.

14. Such an approach to improve response rates in social surveys does not appear to have been tried in African countries though similar attempts have been made to introduce monetary incentives to income and expenditure surveys by giving a small allowance to householders to enable them to keep accounts of household budgets. The African emphasis has been more on paying the interviewer an attractive honararium to enable him to perform effectively.

15. Reference has already been made to the problem of sampling for rare populations. Special sampling schemes have to be drawn up to ensure that reliable estimates are derived from these surveys. In connexion with surveys with the purpose of estimating the blind population, the areas of concentration of the blind have to be determined. In the case of river blindness, the endemic areas are usually well known. Unless varying sampling fractions are used to ensure the proper representation of the blind in the sampled population, the rare population would not be appropriately estimated and its characteristics fully described. The normal approach is to stratify according to proportions of the rare trait and sample more heavily those strata which contain large proportions of the trait. Sampling of rare populations is described in most standard sampling textbooks. 6/ 7/ Leslie Kish considers the following techniques for finding a solution to the problem of sampling for rare traits in the population which could otherwise prove expensive: multipurpose samples, cumulation, large clusters, controlled selection, disproportionate stratified sampling, two-phase sampling, batch testing, and special lists. In considering these approaches, one should bear in mind Cochran's advice that "simple random sampling, or any method of sampling that is adapted for general purposes, is an expensive method of estimating the total number of units of a scarce type".

16. In connexion with the last point, it is recalled that in a number of countries certain areas or categories of the population have been excluded from the target population, because it is argued that they are difficult and expensive to cover. Included in these categories are nomads, remote area dwellers and sparsely populated areas. What is sometimes forgotten is that because of the very nature of these groups, they usually constitute target groups for social concern and action. Thus data on them is very important for monitoring and measuring progress. Suitable sampling techniques which are cost-effective should therefore be evolved for them.

6/ Cochran, W.G. Sampling Techniques, John Wiley and Sons, Inc. New York, 1964.

7/ Kish, Leslie. Survey Sampling. John Wiley and Sons, Inc. New York, 1967.

Conclusion

17. It will be noted from most of the project proposals so far formulated for AHSCP that modules on social conditions feature prominently. For example, Botswana plans to administer a module on fertility, mortality and family planning including questions on social perspectives in 1985. Kenya's programme includes a number of modules which fall under the category of social surveys, e.g., literacy survey (1980), handicapped survey (1981), surveys of health and nutrition and social indicators (1982).

18. It will also be recalled that in the ECA document on Household Data Requirements (E/CN.14/SM/22), 8/ the core items include a number of social aspects. For easy reference the core items are reproduced below:

(a) Community-level variables

- Main source of and distance from water supply
- Distance to and nature of nearest health facility
- Availability of primary school in locality
- Distance from regional capital
- Availability of police post in locality
- Availability of post office in locality
- Distance from nearest market
- Availability of community centre
- Availability of village radio
- Availability of agricultural co-operative/farmer's association
- Distance from nearest motor road
- Cropping pattern or main ecological characteristics

(b) Household as a unit

- Number of households sharing housing unit
- Number of rooms occupied
- Persons per room ration (derived)
- Size of household (derived)
- Total income from farming, employment, trading and other sources
- Numbers and distances of agricultural holdings owned or operated
- Ownership of productive assets

8/ UN Economic Commission for Africa. Household Data Requirements E/CN.14/SM/22. Paper presented to Working Group on Organization, Content and Methodology of Household Surveys. Addis Abata, 15-19 October 1979.

(c) For each household member and visitor

Name

Age

Sex

Relationship to head of household

Residence status

Birthplace

Nationality

School attendance

Highest grade attained

Literacy

Activity status

Occupation

Industry

Employment status

19. There is thus ample scope for the exploitation of data from each round for the study of some aspects of social conditions. The analysis of this information taken in conjunction with data through the use of other modules will present planners with the information needed for the formulation of development plans.

HEALTH

1. Each country whatever its stage of social and economic development, needs quantitative information on the health status of its population, the utilization of health services and socio-economic and environmental factors, affecting health. This information is a basic requirement for planning, management, and evaluation of health services in the country and for monitoring changes in the health status of the population. Depending on its intended uses, this information may be needed at the national, regional, or local level. In some countries some of this information may be routinely generated by the health service. In others little information may be produced by these services, or the health services may be so rudimentary as to provide no opportunity for the generation of such data. In addition, even with the most complete and efficient organization of routine administrative statistics in the health services these are usually important gaps in the information to be filled. For example, health services can usually only generate data referring to users of the health system. Information on non-users, and in particular on their characteristics and reasons for non-utilization can only be derived from other sources.

2. While information that is generated by the health system, or is otherwise collected under the auspices of health system, may serve a large part of the health and other information needs of the country it is often necessary to utilize information from other sources as well. Such factors influencing health and health planning as socio-economic variables (age-sex distributions, nutritional status, employment status and household size and composition are only some examples of information necessary for the proper understanding of health. While one might consider that health information could be fully integrated into a multi-sectoral information system, this chapter will concentrate on those issues which deal with data directly concerning the health sector. Where information from other sectors enters into consideration, reference will be made to the appropriate chapter of the Handbook. The usefulness of national multi-subject household surveys for providing data on health when substantial proportions of the needed health-related information must come from non-health fields should be kept in mind. Specific applications of the multi-subject approach will be considered in the following sections.

3. In order to provide a framework for dealing with the issues of health data needs and health data sources and to examine the place of household surveys in providing for these needs, the concept of a health information system will be introduced. Since the purpose of a health information system is to aid in planning and management of the health services required by the population, it provides guidelines for the population-based data needed for these activities. It also points up the need for population based, problem oriented data, defined in operational terms, and directed to the appropriate decision levels of the health system.

---

* This Annex is extracted from the revised Handbook of Household Surveys, DP/UN/INT-79-020/2/Add.1.

A. Conceptual framework

1. Health information systems

4. A health information system can be defined as a mechanism for the collection, processing, interpretation, and dissemination to the appropriate users of information required for the planning, organization, operation, and evaluation of health services and for research and training purposes. In such a system emphasis is placed in the collection of data on a timely fashion to serve specific needs at local, regional or national levels. The collection and dissemination of the data are seen as being task or action oriented. For example, information for health planning at the national level is of a different nature and is required at different time intervals than is the information required for management of local health services.

5. The information needs of a health system depend in part upon the elaboration and formal complexity of the system. If health planning is carried out at the national level there are likely to be demands for greater amounts of data on long-term potential health requirements (for example, future demands for services, future supply of health manpower, changes in population structure, etc.) than if no planning were carried out or restricted to regional or local levels, or if decisions taken for the short term as part of the management process. At present there is no country in which a full health information system as defined above is presently functioning. However, many countries are increasingly making use of the concept as a means of rationalizing the functioning of their health statistical services.

6. There are six basic constituents of a health information system which are seen as important in meeting the needs of planners and managers of health services and to aid in health service research and training.

These are (1.4):

(i) Indicators of the state of health of the population, such as mortality, morbidity and disability data in varying detail, and various physiological and psychological parameters;

(ii) Measurement of the need for utilization of health services - the use of hospitals, maternal and child health care centres, primary health care centres, etc.

(iii) Statistics on available resources, such as health manpower and its distribution, availability, accessibility and acceptability of available health services and facilities, and financial resources;

(iv) Social, demographic and economic data on the population;

(v) Environmental data, for example on natural risks and dangers created by pollution;

(vi) Outcomes of various preventive and curative measures.

7. Even in the most highly developed information systems there are large gaps in the availability and quality of data relating to these six elements. Much data may be collected from routine sources designed for other purposes, but very often they are either not complete enough or inappropriate for services evaluation and planning. Thus, the collection of mortality statistics may be mostly oriented to legal requirements and the cause of death recorded may be problematic. Hospital statistics seldom include information on ambulatory (outpatient) care. Cause-specific morbidity data are often non-comparable because of lack of standardization of diagnoses. Data from the records of health centres, dispensaries, or private physicians are often inaccessible and very sketchy. Moreover, in many developing countries even the most rudimentary sources of routine health information are lacking and in some, even the basic socio-demographic data on population such as ordinarily are obtained from a national census are not available. The following discussion will examine in greater detail the possible role of the household survey in general, and the national multi-subject household survey programme in particular, in filling the basic data needs of a health information system.

2. Data needs for a health information system

8. One of the functions of a health statistics service is to measure the extent, distribution and nature of ill health, to measure the way in which ill health is translated into demands for health services, and to provide data on the ways in which health services respond to these demands. The totality of these data may be thought of as serving three separate needs:

(i) For planning health strategies at national or regional levels directed at morbidity (ill health);

(ii) For the management and evaluation of health programs which have been designed to meet goals set in (i);

(iii) For identifying relationships important for research into the etiology of health problems.

9. The type of data needed, its sources, and the uses to which it may be put depend upon which of these three uses are to be served. The needs for planning, management, and research may be met by different means and the multi-subject household survey has a role of varying importance in meeting the data needs for each. Each will, therefore, be dealt with in turn, identifying first the data needs of each group, assessing the overlap in data needs where they occur, and finally indicating the way in which the separate and common needs can be met from various sources in terms of routinely collected health data and in terms of *ad hoc* or continuing sample surveys of varying degrees of complexity.

(a) Needs for health planning

10. While it is necessary for health planning to be based on a broad range of properly collected, analysed and interpreted health data, the need for accurate and detailed morbidity data is not as great as that of health managers or health researchers. Those indices of indicators which are needed must, however, be problem oriented and population based. If in addition they can be related to persons and events and are capable of being cross-classified according to demographic and socio-economic characteristics, other social indicators and place of residence, their usefulness to the health planner will be greatey enhanced (1.42).

11. Generally speaking, health planners are interested in measuring those morbid conditions which can be related to specific populations, correlated with other characteristics of the population and the environment, and which translate into needs and demands for health services. The exact diagnostic categories of disease need not be established precisely. Instead emphasis is placed on the impact that morbidity will have on the health needs and behaviour of the population. Thus, data on functional aspects of health, the perceived or real needs of the population for health services based upon disfunction and the socio-economic consequences of disfunctions are the health planners primary needs for morbidity data. Data for morbidity from various nosological categories are (as are cause-specific mortality rates) also of importance in defining the major health problems of a country but for planning purposes the desired accuracy of the diagnosis is much less that that needed for health management and research. In some cases a ranking of most important morbidity and/or mortality causes will suffice in order to establish priorities for resource allocations.

12. Data on utilization of health services are necessary for planning in order to establish where and to what extent gaps in unmet needs and/or demands exist which should be filled by the future provision of health services. For planning purposes it is usually necessary to have knowledge of the characteristics of both utilizers and non-utilizers and to be able to relate utilization patterns to health needs, both perceived and medically verified.

(b) Need for health management

13. Health services management generally requires information for three purposes (2.34). First, data are needed for monitoring the functioning of health services. This primarily involves defining the goals of the health system in terms of resources committed and results forthcoming and seeing what is being done, to whom it is being done and at what cost in resources.

14. Second, data are needed to evaluate the functioning of the health system by a regular and systematic comparison of the care provided by the system with the needs or demands of the population being served. Finally, research is required, given the assumption that the delivery of health care can be improved and ways should be sought by which improvement might be brought about.

15. Both monitoring and evaluation may be thought of as continuing activities built into the health system, while research is more of an ad hoc response to the specific needs, interests and capacities of the researchers. The data needs for monitoring are primarily concerned with measuring coverage of the health system, for example the proportion that benefits from a specified health care activity of a total population that should benefit from that activity (2.93). Knowledge of morbidity patterns is often necessary to evaluate the denominator of this rate since the denominator group consists of persons having a particular health status, for example, pregnant women who also have iron deficiencies, and at whom a specific activity is aimed.

16. Evaluation, as defined above, involves examining changes in health conditions in relation to those resources brought to bear on the system. An important condition is, of course, change in the level of the health status of the population as measured by changes in morbidity in specific parts of the population. Here the morbidity information needed is likely to be biomedically oriented, as distinct from the more socially oriented morbidity measures needed by the health planner. Where specific measurements have been made in order to deal with individual health problems, precise measures of the impact of the health intervention measures are needed. Thus a vitamin supplementation program for young children must be evaluated in terms of changes in the specific morbidity conditions which are due to vitamin deficiencies, such as marasmus, Kwashiorkor, Keratomalacia. The health planner in the same situation might be content to examine changes in mortality rates for all children under five years of age over a relatively long period of time. Health services management also needs breakdowns in morbidity rates by other variables which may influence, or be influenced by, the functioning of the health system. Information for administrative subareas will almost certainly be needed, while various socio-economic and demographic correlates will also be important in the interpretation of changes brought about in health services. Some of these data may be routinely generated by the health care system or be otherwise available from population registration schemes, but other variables may only be known and correlated with morbidity on the basis of continuous or repeated studies.

(c) Needs for research and training

17. Research needs for morbidity data on a population basis tend to be based upon precisely defined medical criteria rather than on more readily available indicators. Here the goals tend to be more scientific than administrative, with precisely defined hypotheses to be tested. Even though the questions to be answered will relate to a whole population (as opposed to small clinical studies which are commonly restricted to selected populations) and will be used to answer questions important for management and planning purposes, the emphasis will be on ad hoc, highly focused studies, using precise diagnostic definitions of morbid conditions. In most cases it is unlikely that research needs for morbidity information can be met by data routinely generated by the health system or other sectors. Nor will the data

collected by laymen not especially trained in complex measurement techniques be capable of providing the desired information. In short, specially designed and necessarily more costly population surveys will probably be needed to meet the special data requirements of these research studies, rather than data than can be generated by multi-subject household surveys.

3. Data sources for health information systems

18. There are, generally speaking, two main sources of data available to health information systems:

(a) That routinely generated by the health or other administrative sectors;

(b) Population based data specially collected either within or outside of the health sector.

In the latter category, household interview surveys are included. Routine administrative data will be dealt with briefly below while population based data sources will be more fully considered, particularly where household surveys are involved.

(a) Routine administrative statistics

19. These are data which arise from the functioning of the health sector. They are for the most part based on records kept by the health system itself, for example, utilization of hospital or other medical services, or queries to various entities of the health system on the health status of those persons coming into contact with them, for example, morbidity data derived from hospital inpatient and outpatient records, or records from doctors' practices. Other sources of morbidity data may be registration systems for various diseases where registration is either mandatory or has been set up for other administrative or epidemiological purposes. Data from medical insurance and other social security plans also fall under this heading.

20. Data on health manpower resources and health facilities are usually available from the health system itself or from other legal registration systems. Where there are licensing regulations for certain categories of health manpower, information is recorded for regulatory purposes. For other categories of health manpower listings may not be readily available centrally and sometimes sampling of records located peripherally may be the most efficient way of estimating the size of certain categories and their distributions in the population. Listings from which information can be obtained are in all cases derived from the health provision system rather than from the general population. Whether complete enumerations or samples are used depends upon costs, details of manpower characteristics needed and administrative convenience. In any case these information sources do not involve sampling of the general population and will not be further considered here.

21. With all of the sources mentioned above, there are problems in identifying the population at risk. While these sources do have the advantage of providing continuous data, these data are often related more to utilization of health care facilities than to actual morbidity. Changes in morbidity rates over time are difficult to interpret since they may reflect only changes in utilization of health facilities. Thus, even when the coverage of a record system is virtually complete for a country's population, as is the case for many health insurance plans, the numerator data refer only to those whose morbid condition led them to use the health care facilities covered by the health insurance scheme. Similarly most record-based systems do not provide the wealth of detail on other characteristics needed to interpret meaningfully the morbidity rates derived from the system. Data relating to an individual's demographic, socio-economic and household characteristics are often unavailable from administrative sources, or if available may be incomplete or inconsistant with other sources in important respects.

22. When health planning is carried out only on a national level, administratively collected statistics may meet some of the health planning information needs, providing the health sector is generating the necessary data and there is some correspondence at the national level in the generation of complementary data from other sources. At the regional and local levels planning based on these statistics may be confronted with a situation where the statistical systems generating complementary data is incompatille with the system of health statistics, rendering impossible or difficult the cross-classification or comparison of health data with data from other sources. In these cases ad hoc ways of providing comparable health and other related data must be utilized. Sometimes surveys of a carefully defined population are the only solution.

23. Some of these record-based systems may meet the requirements of the health manager since they can use more exact definitions or morbidity, provide this information on a continuous basis, and if care is taken in the interpretation of the morbidity rates derived from such sources enable the evaluation of the effect of health care delivery on the health status of the population. Research needs may also be met by special disease registers set up on the basis of record oriented systems, for example, myocardial infarction registers set up on the basis of hospital records and notification by general practitioners for a defined community.

24. Data on coverage and utilization may be obtained both from the providers of health services and from the consumers. Data on utilization obtained from the providers are usually regarded as more comprehensive of conditions and more valid since they are obtained from an established record keeping system. They have the disadvantage however, of being event-oriented rather than person-oriented, and they require some form of record linkage or other analytical work to establish the utilization patterns of individuals. Moreover, the necessary socio-demographic information on an individual which is

needed to cross-classify utilization patterns with explanatory variables, is usually not available from the health system records. Perhaps the greatest disadvantage of routine health services data from providers is that they do not generate any data on non-utilizers and their reasons for non-utilization. Along these same lines they supply little insight into the dissatisfactions of both utilizers and non-utilizers with the services delivered. They are unable to get at those perceived needs of the population for health care which are not translated into demand, or if translated into demand are not met by the health services because of either non-accessibility or non-acceptability of the care of the population served.

(b) Population-based data

25. These include census data, mass diagnostic and screening surveys, health examination surveys, health interview surveys, and multi-subject household surveys.

(i) Census data

26. Census data, while achieving complete population coverage and being free of sampling error, produce limited information because of the infrequency of the census (usually every 10 years) and the highly restricted number of questions which can be asked on health status by interviewers whose main task is to enumerate the total population. However, some data relating to health coverage might be available from census sources, for example, accessibility can be measured in terms of physical distance of various population subgroups from existing health services, or numbers of persons in health professions can be determined.

(ii) Mass screening

27. Mass diagnostic and screening surveys may give highly detailed information about one or more particular disease entities, but they rarely, if ever, cover the total range of morbidity and are, in addition, quite often restricted to certain well defined population subgroups thought to be at high risk for the diseases being screened. Thus these types of surveys possess the capability of providing detailed data on specific morbidity conditions, cross-classifiable with other characteristics of interest, and which may be linked to a well defined (but often restricted) population group. They therefore provide information on the morbidity status of the population which is of use to health managers for monitoring purposes.

(iii) Health examination surveys

28. In order to get a broader picture of medically defined illness in the population, health examination surveys may be of value. In this case a probability sample of the population is invited to a medical examination and history-taking. In this way, even unrecognized, undiagnosed, and asymptomatic conditions may be uncovered, and

distributions of a variety of physical, anthropometric,physiological and psychological conditions in the population may be described and related to a wide range of socio-economic and demographic data simultaneously collected for each person.

29. Surveys may be conducted continuously on the general population using a standard examination of certain index conditions, or at certain time intervals special conditions might be examined, for example, special organ systems such as the cardiovascular or circulatory system. There is also the possibility of examining special population groups, for example, children or the elderly with specially designed examination procedures at certain of the examination cycles. Data in these examinations may also be collected on medical care sought or received during a fixed time period. This allows for identification of those medically determined health needs which are not currently being met by the health system thereby giving health managers certain possibilities of evaluating coverage of the health system. By establishing a series of norms the health planner has, for example, the possibility of following trends of certain health indicators, such as height for age or mid-arm circumference in children, which are of importance in identifying chronically malnourished populations.

30. A health examination survey is likely to provide the maximum information needed on health system coverage, since it is able to measure the amount of perceived health needs and medically verified health needs of the population which are being met by health services. By including appropriate questions on reasons for non-utilization, a good idea can be gained of how the health services are meeting the needs of the population for health care.

31. The major drawback of health examination surveys is the great cost in terms of trained personnel needed to administer the examination and interpret the findings, and the specially equipped mobile examination and interpret the findings, and the specially equipped mobile examination centres often needed to perform the examinations. This restricts the number of persons on which the examination can be carried out. This means that certain estimates will be subjected to large sampling errors when applied to subpopulations, or when the morbid condition or other characteristics to be estimated occurs relatively rarely in the population being examined.

32. Health examination surveys also share with other survey methods the problem of non-representativeness due to incomplete co-operation of the population sample. Since lack of co-operation is unlikely to be proportionally distributed across the population, biases may occur in morbidity or other estimates. Usually a substantial part of a survey budget must be devoted to minimizing the number of incomplete examinations and/or to independent studies that must be carried out to evaluate the effects of non-response.

(iv) Health interview surveys

33. The health interview survey, which makes use of questionnaires, administered by an interviewer in the home or in some cases mailed to or left with the potential respondent to be mailed back to the sponsoring agency, is one of the most common methods of collecting data on the health status of the population. These surveys may treat specific population groups or deal with restricted health nutritional status of children in rural areas, environmental exposure of certain occupational groups - but the focus here is on large-scale national studies having the following features:

(1) Data are collected on a wide variety of health and health-related topics rather than for a single purpose

(2) The studies are conducted on a national basis on large representative population samples with the intention of providing national and subnational estimates of known precision.

(3) The interview method is used and the interview conducted by lay interviewers. Only very simple objective tests or measurements may be used and professional diagnostic services or devices are not used;

(4) The surveys are continuous or at set intervals but have limited longitudinal capabilities. Even panel surveys rarely follow individuals or households over more than one year. Thus the type of morbidity information the survey is capable of producing is point or period prevalence rather than incidence data over time.

34. The potentialities and limitations of health interview surveys have been documented elsewhere (2.33, 2.96) and may be briefly summarized here as follows:

(1) Broad subject coverage is possible compared to data collected from record based systems, and the data consequently serve a general range of health information needs;

(2) As opposed to data collection systems with specialized objectives based on other than health considerations which therefore cannot be easily modified to provide currently useful data, the health interview survey is a flexible instrument which is responsibe to changing health information needs. The data collected are determined (within the limits of resources) by current information needs and not by extraneous administrative conditions.

(3) Representativeness of morbidity is assured since, if a properly designed and carried out probability sample of the population is used, the appropriate denominator data are automatically generated for the calculation of various rates. Data on morbid conditions collected from hospital records or medical practitioners can seldom provide denominator data, thus restricting the usefulness of the data for the evaluation of health status;

(4) The individual is the focus of the enquiry, allowing for the collection of health or health-related series not available from record oriented systems. In particular, data are also available for those individuals who do not come into contact with the health care system. This may be the only source of information on the characteristics and health status of non-utilizers of health services. Even when data can be collected from various record oriented sources, it is usually not feasible to link these data for individuals in a way permitting cross-classification

(5) The major disadvantage of the health interview survey is the nature of data that may be collected by the interview method using lay interviewers. These may be referred to as response errors, and should be distinguished from sampling errors, which may be substantial but are estimable in advance in most cases and unlike response errors have consequences which may be objectively evaluated. Response errors will be discussed in more detail in section I below.

(6) Morbidity as measured from health interview surveys is necessarily that perceived by the respondent and is not necessarily that which has been medically verified as arriving from a pathological conditions to which a diagnostic entity can be attached. Emphasis is based on those conditions which have caused the respondent over a short or long period to reduce temporarily or permanently his habitual activities or to seek aid from the health care system, or at least to express a need for such aid. This form of measurement of health status would appear to best suit the information needs of health planners since it provides prevalence data which may be classified in broad categories such as acute conditions, minor and major chronic conditions, impairments, injuries and disability.

35. The household health interview survey is likely to be the most feasible instrument for collecting data on utilization from the population point of view. Compared with the health examination survey, it has the disadvantage of dealing with the perceived needs of the population for health services, rather than medically determined and medically verified needs, but this might not be truly a disadvantage since it is the perceived needs of the population that determine pressure on health services. At the same time the interview survey is capable of assessing the attitudes of both those who utilize and

those who do not utilize certain types of health services of measuring what the respondents feel are the barriers to use and what are their major satisfactions and dissatisfactions with the structure and functioning of the health delivery system. This type of data, when integrated with data provided by the health services themselves, can give a more complete picture of utilization and the potentials for improvement in coverage and efficiency of the service.

36. It is important in considering the feasibility of a survey whether utilization data are being gathered as part of the planning and evaluation aspects of health management or whether they are being gathered as part of a research study. The information derived from a one-time research study where consumer needs are examined in depth may justify the cost of the household survey in that a large amount of in-depth data can be collected by means of a detailed questionnaire. In this case valuable information on health consumers' needs and attitudes can be gathered, perhaps in advance of and then following some innovation in health planning (for example, the establishment of a new type of service) or some change in health management (for example, new staffing patterns for health clinics), which will aid in the evaluation of such innovations. It should not be expected that such special purpose household surveys can be mounted routinely because of the dedicated nature of the information being collected and the detail and length of the questionnaire needed to answer the questions posed by new innovations in health care delivery. Up to the present, most health surveys have been of the one-time type designed to answer specific questions (not only on the effects of changes in delivery of health care, utilization and coverage, but also on the effects of health intervention programs on health status) at a cost which could not be justified if such surveys were to be carried out routinely to provide surveillance or monitor activities of the health system.

37. If carried out for the purpose of monitoring and evaluating utilization of health services at the national level, health topics may be incorporated into existing continuing multi-subject national household surveys in order to achieve the economies inherent in this type of investigative tool. It should be realized, however, that all the detailed information on needs for and utilization of all types of services and on the attitudes of consumers and non-consumers cannot be incorporated in a questionnaire devoted to several other subject-matter fields as well. The fact, however, that information on other topics is included with the questions on health services need and utilization may provide other possibilities for analysis by cross-classifying health needs and utilization data with demographic, socio-economic and household data not ordinarily collected when the survey is sponsored only by the health sector.

38. Data on the use in the household of substances which may be harmful to health may be collected by means of a household survey provided the use is a legal and socially acceptable one and one on which the respondent is willing to report. Use should also be widespread

enough in the general population for the household survey to be an efficient method of collecting the data. Data on exposure to or use of noxious substances connected with the work experience of individuals might be better obtained by sampling places of work. If the use of noxious substances is restricted to certain segments of the population, then sampling might better be carried out where these individuals congregate rather than using the household as the sampling unit. For example, drug usage surveys are often carried out on university or secondary school students or young military personnel using samples derived from rosters of these individuals. Even though large segments of the population will be missed, few users might be missed since the sampling has been concentrated in those parts of the population at high risk.

39. Drug and alcohol usage questions can also have a detrimental effect on multi-subject surveys since they may create opposition by the respondent to answering other questions in the questionnaire. It is also unlikely that a proxy respondent can give reliable answers for another person for these types of questions. While questions on drug or alcohol usage might be acceptable in a survey uniquely concerned with health questions and sponsored by a health agency, it is less likely they will be welcomed where the sponsorship of the survey is some other governmental agency. On the other hand, such subjects as tobacco usage or potential exposure to various noxious agents such as polychlorinated biphenyls (PCB) in the household environment, usually meet with little respondent resistance.

40. While data on environmental hazards which affect large segments of the population, such as air or noise pollution, can be obtained individually from each household, ambiant levels can be obtained directly by sampling the environment itself. However, the consequences of various levels of environmental pollution (respiratory difficulties from air pollution, psychosomatic problems from noise, etc.) must be measured by means of a household survey. If a household survey is used, then only perceived conditions due to environmental pollution will be recorded, but rough measures of consequences might be obtained from households subject to different levels of environmental pollution.

B. Selected issues concerning health topics in household surveys

1. Definition of health status in household surveys

(a) General considerations

41. The Constitution of the World Health Organization states in its preamble: Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity (2.90). In view of this definition, it is perhaps ironic in considering the health status of a population that it is departures from health, that is, illness or morbidity, rather than the positive aspects of health that are almost always measured. This is not only because the concept

of health is itself usually defined by the absence of illness but also because it is simpler to measure morbidity than it is to operationalize for measurement purposes the more complex positive concept of health.

42. An illness or morbid condition may be defined as a departure, subjective or objective, from a state of physical or mental well-being, as a result of either disease or injury. An individual may or may not be aware of this condition, and the condition may consist of active or progressive disease or of chronic or permanent defects resulting from disease, injury, or congential malformation. If the presence of the morbid condition is determined by means of a health record from a physician or hospital or from a health examination survey a diagnostic category can be attributed to the condition and the severity of the condition can be objectively evaluated.

43. The subject need not have been aware of the condition and the condition need not have had any influence on the subject's activities. For example, a subject, as a result of a visit to a doctor or as the result of a health examination by a qualified technician on a survey basis, may be discovered to be suffering from arterial hypertension. He may not have been aware of the condition and unless the condition was quite severe would not have modified his usual activities because of it.

44. On the other hand, in order for a morbid condition to be reported in a household interview survey, the subject must be aware of the presence of the condition, and it usually must have caused some modification in his daily activities or necessitated a visit to the health services. Thus the type of morbid condition covered by the household interview survey is that which is perceived by the respondent, which has caused him to take some action because of the condition and which may or may not correspond to a medically-verified diagnosis of illness. The subject is often able to identify more or less precisely the diagnostic category in which his morbid condition falls; but this depends upon the level of communication with the physician and whether the subject has been given precise information about his condition.

(1) Acute and chronic conditions

45. Household surveys usually attempt to distinguish between acute and chronic disease conditions.

46. Acute conditions are frequently characterized by sudden onset and are of finite duration. In addition they tend to severely restrict the subject's usual daily activities. These three characteristics lead to a simple and reliable method of estimating prevalence by household survey since the recall period is short, the event has had an impact on the respondent, and the period of illness is usually well-defined with a specific medical diagnosis.

47. In the case of chronic illnesses the onset is usually insidious, with a gradual progression of symptoms or with problems of a more permanent nature resulting as sequels to a series of acute conditions. Daily activities may or may not be restricted during any given period although there is usually a more general series of limitations of activities. Medical diagnosis may be more difficult, treatment less sure and demands on health services, while perhaps less intense over a short period, will also be of long duration. It is the consequences of chronic conditions that usually lead to a need to measure disabilities.

(c) The person and condition approach

48. Two main strategies for collecting morbidity data appear to be currently used in household interview surveys. These are described below.

(i) The person method for acute conditions

49. The presence of the morbid event is measured by queries as to reduction in a respondent's usual activities due to a health problem during a certain reference period (for example, two weeks before the interview), or about contacts with the health services during another reference period (for example, 12 months before the interview). These are usually referred to as acute episodes of illness or impairment. In addition the duration of the episode is queried, and very often the cause of the episode is also asked. The cause is elicited either by asking the respondent to describe the condition in his own terms or with the aid of a checklist of conditions. The checklist may or may not correspond to standard diagnostic categories found in the International Classification of Diseases (1.43) and is usually phrased in ways meaningful to the respondent. The main purpose of the checklist is not to estimate prevalence of the condttion but to act as a recall aid on the range of illnesses being considered, some of which may otherwise be overlooked by the respondent.

(ii) The condition method for acute conditions

50. This approach differs from the person approach in that initially only those health conditions are queried that caused the person to take action or to change his usual activities. The condition approach asks about all illnesses during the reference period whether or not the illness had any modifying effect on the respondent's health behaviour. Whether these conditions were acute or chronic is established later as is the action (if any) that is taken by the respondent because of the condition (restriction of activity, consultation with a health service). The condition approach includes minor illnesses that may have required no action of any kind on the part of the respondent and potentially serious conditions which do not yet require action.

51. It has been shown that the person approach by restricting attention to those conditions which have an impact on persons has the following characteristics:

(a) It produces an emphasis on health problems of importance to the population;

(b) It reduces the amount of interviewing time needed to elicit this information;

(c) It reduces response errors found to be present when reporting events of minor importance to the respondent. (2.34)

The same advantages also apply to the use of the person approach to estimating the prevalence of chronic conditions where initially limitation in activities over a long reference period (usually 12 months) is recorded. Only for those persons reporting some limitation of activities during the reference period is the duration and cause queried.

52. The person approach to recording morbidity emphasizes the ability of the household interview survey to obtain health status data which emphasizes those conditions which have an impact on the person's usual activities, cause him to seek medical care, and therefore have social and economic consequences for the person, the community and for the health delivery system. The approach does not provide an exact biomedical description of the health of the community in terms of diagnostic categories and is of little value for obtaining cause-specific prevalence data.

(d) The measurement of disability

53. The measurement of disability by means of household surveys raises some special problems not the least of which relates to terminology currently used. Such terms as disability, impairment, disfunction, handicap, disorder, infirmity, etc., are often used interchangeably to express the same condition, while in some studies a distinction is drawn between these conditions. Thus it is not always clear whether a household survey reporting an estimate of prevalence of handicaps in the population refers to the same events as another household survey reporting an estimate of the prevalence of disability. Only examination of the questions posed and the probes used to elicit responses can resolve the question of comparability.

54. An attempt has been made to bring some order to this problem of terminology and it is strongly recommended that when attempting to measure the consequences of disease using household surveys the classification scheme outlined in the WHO manual should be taken as the standard (1.44).

55. The consequences of disease, particularly chronic conditions, may be presented schematically as follows: disease - impairement -- disability - handicap. The arrows represent the logical order of the event in a sequence of consequences arising from the disease.

56. The definitions of the terms in (1.44) are as follows

(1) An impairment is any loss or abnormality of psychological, physiological or anatomical structure or function. This includes what are generally called 'disorders' and in addition losses, such as loss of a limb. An impairement may be temporary or permanent. Impairments are disturbances at the organ level.

(2) A disability is any restriction or lack (resulting from an impairment) of ability to perform an activity in the manner or within the range considered normal for a human being. Disabilities involve activities and behaviours in personal care, in the performance of usual activities of daily living, and in locomotor activities. Disabilities reflect the awareness at the level of the person of an impairment. A person can be impaired but not necessarily disabled if the impairment does not interfere with the person's ability to perform according to the definition.

(3) A handicap is a disadvantage for a given individual resulting from an impairment or a disability that limits or prevents the fulfilment of a role that is normal (depending on age, sex and social and cultural factors) for that individual. It is characterized by the differences between an individual's performance and status and the expectations of the individual himself or of the particular group of which he is a member. Handicaps may be thought of as the socialization of disabilities and impairments. A person may suffer an impairment and may be disabled, but may or may not be handicapped.

57. A household survey using lay interviewers may measure disabilities and handicaps, since both of these can be arrived at by questions dealing with limitation of activities over a long period of time. For example, the United Kingdom General Household Survey (1980 version) asks the following questions (2.47)

"Do you have any long standing illness, disability or infirmity? By long standing I mean anything that has troubled you over a period of time or that is likely to affect you over a period of time."

If yes

(a) "What is the matter with you?"
(NO PROBES)

(b) "Does this illness or disability limit your activities in any way?"

The answer to part (a) of the question determines the nature of the impairment as perceived by the respondent and stated in his own words, while part (b) determines whether this impairment is a disability and perhaps also a handicap.

58. The distinction between a disability and a handicap may be important if one of the aims of the survey is to assess unmet needs and demands for health, since a disabled but not handicapped person may not have any need for rehabilitative services if he is performing to his and his group's expectations. Further probes with the type of question illustrated above are likely necessary to distinguish between disabled and handicapped status. If the survey is dedicated only to the causes and prevalence of impairments, disabilities and handicaps, then a large number of supplementary questions on the exact nature of the condition, its antecedents, its severity, and its consequences should also be considered. Few of these could be considered in the context of a general household survey, however, with general planning or management purposes. A specialized survey is likely to be needed to provide data for research or for evaluation of specific programmes. In these instances a specially trained staff can be employed to evaluate the history of disease leading to the event and to grade the severity of the event itself.

(e) Lay reporting of mortality and morbidity

59. In many areas of the world, a large proportion of the population has no access to health care provided by medically qualified personnel. This is particularly true of the developing countries. In these areas, health care is often provided by lay or paramedical personnel and is based on traditional methods or elementary medical training. The health information needed to indicate the existence of health problems or to facilitate the management of primary health care systems needs must to a large extent be provided by these same personnel. The term "lay reporting" has come to be used to describe methods of achieving the provision of this information (1.46).

60. The system of lay reporting is based on a series of assumptions

(1) A person familiar with a local community and who possesses administrative or moral authority, such as a village chief or an auxiliary health worker, may learn of any death occurring in the community;

(2) An auxiliary health worker, adequately trained, may be able to record symptoms and associations of symptoms as expressed by the patients;

(3) The circumstances of death and the manifestations of diseases as recorded by the layman or paramedical personnel, may be classified by a suitable and simple method and interpreted by a qualified health worker in such a way as to prove useful for practical health actions.

61. Several requirements must be met: (1) Each stage in data collection and in particular the means for the identification of health problems and recording of data should be so simple and unambiguous that the least skilled grade of staff can perform efficiently. (2) The system should be practical under rural conditions, without the need for costly facilities or additional health personnel, and should fit perfectly within the administrative structure of the health service; (3) Acceptable standards of accuracy should be achieved and maintained systematically during the operation of the system.

2. Response errors in household surveys

62. Response errors arise from various sources, most of which are found in the household interview situation. 1/ These are described below.

(a) Defective recall of morbid events

63. The problem of recollection is a complex one since some types of events, for example, those requiring hospitalization or those involving considerable trauma, have recall periods much longer than events having less serious health consequences. In order to record as many events as possible and thus reduce sampling errors, there is a temptation to utilize long periods of time (say a one-month or a one-year period). It has been shown, however, that recall falls off rapidly even over a 4-week period for those morbid conditions for which no medical care or restriction of activity is required (2.35). Over the same 4-week period, there is less evidence of memory loss for those conditions involving medical care and activity restriction.

64. Another problem with memory loss is that certain cultures use different frameworks to measure time, and periods involving weeks or months may have little meaning for these respondents. It is usually necessary to adopt a time frame which is important for a given culture providing that the events constituting the tome frame (a period of harvest, a religious festival, etc.) are not too far removed from the data of the interview.

65. It thus appears that each class of morbid events has its own best recall period. Thus, the General Household Survey of Great Britain (1980) uses the following recall periods (2.47):

(i) Any illness or injury involving restriction of activities - 2 weeks

(ii) Consultation with doctor - 2 weeks

(iii) Use of health or welfare services - one month

1/ Experience with response errors in surveys in the field of health in the U.S.A. is reviewed in (2.1).

(iv) Use of casuality ward, outpatient or inpatient services of hospital - 3 months;

(v) Chronic limitation of activities - unlimited recall period.

(b) Non-familiarity with the nature of the morbid condition

66. Very often the respondent is not familiar with the medical nature of the condition from which he is suffering or has suffered. This may be because he has not consulted the medical services or because an exact diagnosis has not been communicated to him by the health services. Even when a person is advised of a condition, it is not described in sufficient detail to be classifiable into an exact diagnostic category. For example, the subject may be told he has a "heart condition" instead of "acute myocardial infraction". Many physicians do not inform persons of the exact nature of certain cancers by may describe the condition in less threatening terms. Many morbid conditions are in addition non-symptomatic, for example, hypertension and anaemia so that the respondent may not be aware of their existence. The consequence of such lack of familiarity is both underreporting and misclassification of conditions.

(c) Reluctance to report

67. Certain conditions may be underreported because of reluctance of the interviewee to discuss them with a lay interviewer. Examples of this are certain mental illnesses, venereal diseases, genito-urinary conditions and alcoholism, for which a certain amount of shame (or societal disapproval) is felt. Reluctance to report these conditions may vary from one social or demographic group to another, vary among different cultures, or change over time. Thus the result of unwillingness to report conditions does not necessarily result in uniform underreporting of these kinds of conditions, but might result in artifactual differences being discovered among different sub-groups of the population or between different times for the same population.

(d) Proxy respondents

68. In most household interview surveys, each individual residing in the selected household is a basic unit on which health status information is required. Since not all household members are usually present when the interviewer calls, a decision must be made either to call back at the household until all household members can be interviewed or to accept an adult member of the household found at the initial visit to serve as a proxy respondent for all other household members. It is the usual procedure for a parent to serve as proxy respondent for children. It has been estimated in the U.S. National Health Interview Survey that direct field costs of the survey are approximately 25 per cent higher when each household member is interviewed, but that the estimates of prevalence of certain conditions tend to be underestimated when proxy respondents are accepted. It has also been found that minor non-chronic conditions tend to be underestimated more than major non-chronic condition conditions or chronic conditions

(e) Probe question

69. In order to refresh the respondent's memory of past morbid events or to ensure that the respondent fully understands the full range of morbidity being queried by the survey, a series of probe questions, which may also include the presentation to the respondent of a check-list of conditions, is usually administered. It has been shown that a final probe with a checklist can increase the number of all chronic conditions found by over 50 per cent, and for those chronic conditions not involving medical care or restricted activity by 67 per cent. Thus two surveys querying the same set of conditions but using different systems of probes can arrive at quite different prevalence rates for the some morbidity condition (2.85).

(f) Differential biases

70. A further type of response bias has been noted in health interview surveys involving diary keeping by the respondents on their health conditions (2.42). It has been found that higher socio-economic status, higher income, and urban households tend to report a higher prevalence of morbid conditions. This is thought to be partly due to their greater experience with medical services greater sensitivity to disease conditions and greater aptitude for reporting these conditions. Differences tend to be greatest in reporting minor and obscure diseases and injuries.

3. Needs, demands and utilization of health services

71. It has been noted that morbidity either as perceived by the subject (see sub-section 1 above) or as verified by a medical diagnosis does not completely explain the utilization of health services (2.88a). While perceived morbidity has been shown to be important in the utilization of physician and dental services (2.31), other factors have also been important in explaining utilization. Information on factors which have been designated as 'predisposing' and 'enabling' factors are usually not available from health service records and if available, only for those who have utilized the health service.

72. Predisposing factors are those socio-cultural variables that predispose or deter a person from using health services. These may include demographic characteristics, household and family composition, education, attitudes to health etc. Enabling factors are those that relate to 'cost' in a broad sense, and may include family financial resources, availability and accessibility of health care, health insurance, income security etc. The concept of utilization of health services is that perceived morbidity by an individual represents the individual need for health services, the series of predisposing factors determine whether the need is translated into a demand for services, and the enabling factors determine whether the individual uses the health services. There is of course an interaction with factors which are characteristics of the health system itself that is, the existence of the facilities for which a demand is manifested, and information on this latter aspect is best collected from the health system itself.

73. Data on these factors are well suited to collection through household inquiries, since the household interview survey allows for putting together household, family and individual data, including perceived individual morbidity in a way which can account for utilization patterns and for reasons for non-utilization.

C. National experience

74. While many countries carry out ad hoc household surveys on health and health-related topics at either the national or sub-national level, few use this instrument in a continuing fashion for the routine collection of data for health planning (2.96). In those countries which do conduct such surveys regularly for example Japan and the United States of America, they are dedicated primarily to health data with data from other fields included only inasmuch as they are thought to be essential to the interpretation of the health interview data. In the United States of America, the Health Interview Survey is one of a series of national health surveys that also includes a health and nutrition examination survey and a national inventory of hospital and institutions which allows for the sampling of medical and institutional records.

75. In addition, the Federal Republic of Germany and the United Kingdom carry out multi-subject household surveys on a continuing basis, that in the Federal Republic of Germany every two years as the micro-census programme and that in the United Kingdom annually as the General Household Survey (GHS). In addition to data on health and health services utilization, the GHS collects data on employment, housing, income, education, and special topics which change on a rotational basis. The micro-census in the Federal Republic of Germany uses an area random sample based on the previous population and occupation census. The sampling fraction is 1 per cent. Data on acute and chronic diseases, injuries due to accident, and disabilities are recorded. In addition, information on various environmental pollution risk factors and smoking are included as well as utilization of health services.

76. The GHS covers a probability sample of about 15 000 households each year. The major health topics are activity limitation due to sickness, consultations with general practitioners, use of health services and visits to hospitals.

77. The specific content of questionnaires can differ greatly both in the way in which data are recorded and in the different items which may be included to get at morbidity and health services utilization. The content of five national health interview surveys conducted during the 1950s is compared in [redacted] tables 1 and 2 (from 2.33). While the specific questions have likely changed in those three countries where the surveys are still conducted on a continuous basis (Japan, United kingdom, United States of America) the tables do illustrate the different specific items that have been used by each country to get

morbidity and health care utilization information using the same general conceptual framework. Thus, all national household interview surveys tended to use the same general concepts of morbidity, querying prevalence of acute and chronic conditions, impairment and injuries due to accident and disabilities. Table 2 illustrates differences in information actually obtained on various aspects of these morbid conditions.

78. Differences in the recall periods for various classes of morbidity and utilization are shown in [redacted] table 4. Again wide divergencies in recall periods are noted even when the same general question is asked Since there are no established standards for health interview questions, it is impossible to give guidelines for optimal forms of questions, recall periods, probes, etc., that might be suitable for all national household surveys. Improvements in data collection from national household health interview surveys carried out in the United States of America which have been documented indicate that changing the focus of some questions and including expanded checklists of chronic conditions leads to improvements in the underreporting of certain chronic conditions and long-term disabilities (2.00).

79. It was found that the percentage of persons reporting chronic illness in the General Household Survey of Great Britain increased from 25 per cent to 56 per cent in interviewed males and from 26 per cent to 70 per cent in interviewed females between the 1976 and 1977 surveys, when the format of the question was changed in 1977 to include a checklist of conditions (2.47). It is thought that the increases in prevalence of chronic illness was in large part due to the inclusion of conditions in the checklist (for example, backache, varicose veins) which the respondents to the 1976 format did not feel to be serious enough to volunteer. Several of the items on the checklist concerned women more than men (for example, varicose veins, menstrual and menopausal difficulties) which may account in part for the greater increase in prevalence among women than among men. It is not entirely clear whether the 1976 format yielded an underestimate or whether the 1977 format should be considered as giving an overestimate of chronic illness. The 1979 and 1980 formats have dropped the checklist and have returned to reporting separately those chronic conditions which limit activities and which are closer to the 'person' approach than to the 'condition" approach described in section F above. Whether these improvements would be seen by applying the same approaches in different national contexts and whether these techniques would be feasible in multi-subject household surveys where relatively little time can be spent in the interview administering detailed questions and probes on health are questions for further research in health interview survey methodology.

80. Questions and interview techniques appropriate to a particular health system in a particular cultural setting may not be appropriate when dealing with other health systems in other cultures. An example of an international comparative study on health care carried out in 12 centres in seven countries using a household interview survey is

the WHO/International Collaborative Study of Medical Care Utilization (2.31). This study demonstrated that it is possible to obtain locally valid and cross-nationally comparable data from a common household interview survey questionnaire on such topics as psychosocial and demographic characteristics, levels of health and perceived morbidity, and detailed use of various health services using a deteiled and carefully planned questionnaire (302 questions). However, all but one of the centres (Buenos Aires) was either from Europe or North America, so that the range of cultural variability was not as great as it might have been if Asian or African populations had been included in the sutdy. All seven countries were industralized, with well developed health care systems already in operation so that the measurement errors due to different conceptual meanings of certain questions and their responses were minimized.

D. Data requirements and data tabulations for health topics

1. Data requirements

81. As has already been emphasized in section C above, there appears to be no universally acceptable question format or set of health questions that can be collected by all countries by means of national multi-subject household surveys. Similarly, data requirements for special health surveys carried out for research purposes or for evaluation of specific programmes must follow specific formats depending on the aims of each study, the research design employed, and the resources available. Gneral guidelines on format and content cannot be given in these cases without taking account of these considerations.

82. There does appear to be, however, a core content of health items for general multi-subject household surveys which can be collected by means of interviews, using lay interviewers. The exact format of the questions cannot in general be given since this will depend upon conditions particular to a country, for example, if a checklist of conditions is presented to the respondent as a recall aid, or as an indication of the range of health conditions being considered, the content of the checklist would be quite different in an industrialized country with a temperate climate than in a developing country with a tropical climate.

83. A report of the United Nations Economic Commission for Africa has suggested a list of health and health-related items which could be included as a special health module in addition to core questions which are included in all survey rounds. The health module would alternate in certain annual rounds with other modules such as household income, consumption and expenditure, detailed demographic characteristics, employment, etc. Items are asked for each household member. These items may be grouped as follows:

(a) Health status:

(1) Illness during last 14 days:
nature of illness
duration of illness during last 14 days
number of days of restricted usual activity:

(2) Injury during last 14 days:
nature of injury
external cause of injury
number of days of restricted usual activity due to injury:

(3) Impairments (no time limit):
nature:

(b) Utilization of health facilities:

(1) Consultation for health care:
number of visits during the last 14 days by place of visit
type of service rendered
date of last visit (if before 14 days) stating the place of visit:

(2) Hospitalization during the last 12 months:
number of admissions
duration of hospitalizations

(3) Expenditures
drugs and medical appliances prescribed by a doctor or those under his supervision
hospital expenses
doctor's fees
other (specify):

(c) Other items related to health:

(1) Nutrition this includes anthropometric measurements on young children and mothers food consumption history breastfeeding and weaning information

(2) Housing and water supply: include items on sanitary facilities in the home, waste disposal facilities, and source of water supply.

84. To this list of items which is seen as meeting health information needs of developing countries in Africa, could also be added.

(a) Further items on chronic illness (with a recall period of 12 months) which may or may not limit activities. A checklist of items appropriate to the national health situation should be presented in order to remind the respondent of the types of conditions being queried. Where a subject cannot read, the interviewer will have read the list to the subject. Each condition should be described in locally understood terms.

(b) An extension on the item of impairments in order to determine the extent of limitation of activities and their consequences for the subject's particular situation, that is, the need to take special medicines or use special aids, special diets, limit certain habitual activities, etc.

85. All the items listed above could be included as core items for a health component of a national household survey when resources permit the administration and subsequent tabulation of more comprehensive survey content at each survey round. The General Household Survey of Great Britain is an example of a multisubject survey which includes several topics at each survey round, although the specific items constituting a topic may change from round to round to reflect either changing interest or changing concepts relating to that topic.

86. The illustrative listing of items given above should not be confused with the specific question format which might be used in a questionnaire in a particular national context. The way in which the question is asked, the response categories listed and the positioning of the question in the questionnaire, will depend upon specific circumstances, including the training and education of the interviewers cultural and educational characteristics of the population interviewed, and data processing facilities available for tabulations.

2. Tabulations of health data

87. Both health status and health behaviour usually vary by both age and sex. Prevalence of both acute and chronic conditions increases with age, and women usually report more illness and disability than men. Utilization of health services also varies with age, with some facilities, for example, maternal and child-care centres used uniquely by young children of both sexes, or by women of childbearing age. Thus, almost all tabulations involving either health status or utilization are cross-tabulated by age and by sex. The other variable by which many health variables are also classified is socio-economic group. The definition of socio-economic group varies from one culture to another and various proxies such as income, occupation, education, housing, or various combinations of these have been used to describe it. In specific national contexts, other variables may also become important for cross-classification, for example, marital status, ethnic group, family size, rural or urban residence, physical access to health care facilities. If the household survey is dedicated to special health topics, then the modifying variables will depend upon the research aims of the study.

88. As an example of tabulations possible from a multi-subject household survey, 12 tabulations of the 45 presented by the General Household Survey (1977) of Great Britain are given below (2.47):

(1) Persons reporting chronic health problems, by sex and age;

(2) Persons in contact with health services in the year before;

(3) Persons who did something differently from usual in the 14 days before interview because of chronic health problems, by sex and age;

(4) Persons consulting a general practitioner (GP) in the 14 days before interview about chronic health problems, by sex and age;

(5) Persons reporting short-term health problems in the 14 days before interview, by sex and age;

(6) Persons consulting a GP in the 14 days before interview about short-term health problems, by sex and age;

(7) Health problems in the 14 days before interview, by sex and age;

(8) Persons consulting a GP in the 14 days before interview about health problems, by sex and age;

(9) Persons reporting chronic health problems, by sex and socio-economic group;

(10) Persons taking constant care because of chronic health problems, by sex and socio-economic group;

(11) Doctor consultations:

(a) Persons consulting a GP (National Health Service - NHS) in the 14 days before interview, by sex, age, and socio-economic group: rates per 1000;

(b) Average number of GP (NHS) consultations per person per year, by sex, age, and socio-economic group;

(12) Outpatient attendances:

(a) Persons attending "outpatients" in a 3-month reference period, by sex and age: rates per 1000;

(b) Average number of attendances per outpatient in a 3-month reference period, by sex and age.

The major cross-classifications in this context involve age, sex, socio-economic group and marital status.

CONTENT AND RECALL PERIODS IN HEALTH INTERVIEW SURVEYS IN FIVE COUNTRIES

Table 1. Survey content: health services

	Canada	Denmark	England and Wales	Japan	USA
Medical services.....	Yes	No	No	No	Yes
Consultations, total number.	Yes	No	No	No	Yes
Number by place of consultation.....	Yes	No	No	No	Yes
Nature of service received..	Yes	No	No	No	Yes
Place of consultation.....	Yes	No	No	No	Yes
Elapsed time since last visit to a doctor.....	No	No	No	No	Yes
Hospitalization.....	Yes	Yes	Yes	No	Yes
Date of admission.....	No	Yes	No	No	Month only
Name and address of hospital	Yes	Yes	No	No	Yes
Nature of condition.....	Yes	Yes	No	No	Yes
Number of stays, overnight or longer, in:					
Hospital.....	Yes	Yes	Yes (days)	No	Yes
Nursing home.....	No	No	No	No	No
Sanatorium.....	No	No	No	No	No
Operation performed?.....	Yes	Yes	No	No	Yes
Dental services.....	Yes	Yes	Yes	No	Yes
Total number of consultations	Yes	No	Yes	No	Yes
Types of expenditure for health.....	Yes	No	No	Yes	No
Appliances.....	Yes	No	No	No	No
Equipment.....	Yes	No	No	No	No
Health services.....	Yes	No	No	Yes	No
Method of payment.....	Yes	No	No	Yes	No
Medicines.....	Yes	No	No	Yes	No
Doctor-prescribed.....	Yes	No	No	Yes	No
Self-prescribed.....	Yes	No	No	Yes	No
Premiums for pre-paid plans	Yes	No	No	Yes	No
Kind.....	Medical and hospital combined	No	No	Medical, hospital and dental	

Table 2. Survey content: morbidity conditions

	Canada	Denmark	England and Wales	Japan	USA
Duration of illness.....	Yes	Yes	Yes	Yes	Chronic only
Date of onset.....	Yes	Yes	Yes a/	Yes	Yes b/
Date of termination.....	Yes	Yes	Not stated	Yes	No
Days illness during survey period.....	Yes	Yes	Yes	Yes	No
Interference with usual activities.....	Yes	Yes	Yes	Yes	Yes
Days kept indoors.....	Yes	Yes	Yes	No	No
Days in bed.....	Yes, home and hospital	Yes, home and hospital	Yes, home and hospital	Yes, hospital only	Yes, hospital only
Days prevented from working	Yes	Yes	Yes	Yes	Yes
Days prevented from attending school.....	Yes	No	No	Yes	Yes
Source of income if not working.....	No	Yes	No	No	No
Nature (name) of condition	Yes	Yes	Yes	Yes	Yes
Number of conditions.....	Yes	Yes	Yes	Yes	Yes
Outcome.....	No	No	No	Yes	No
Circumstances of accident.. resulting in injury.....	Yes	Yes	No	No	No
Place injury occurred.....	Yes	No	No	No	Yes
Site of condition.....	Yes	Yes	Yes	No	Yes
Treatment for each condition listed.....	Yes	Yes	No	Yes	Yes c/
Medical services.....	Yes	Yes	Yes	No	Yes
Consultations, total number.....	Yes	No	Yes	Yes	No
Number by place of consultation.....	Yes	Yes	Yes	No	No
Date of first consultation.....	Yes	No	No	Yes	No
Date of last consultation.....	No	No	No	Yes	For chronic conditions only
Physician's diagnosis	Yes	Yes	No	Yes	No
Place of consultation	Yes	Yes	Yes	Yes	No

Table 2. (cont'd)

	Canada	Denmark	England and Wales	Japan	USA
Hospitalization.....	Yes	Yes	Yes	Yes	Yes
Number of admissions.....	Yes	Yes	No	Yes	Yes
Date of admission.....	No	Yes	No	Yes	Yes
Date of discharge.....	No	Yes	No	Yes	No
Days hospitalized.....	Yes	Yes	Yes	Yes	Yes
Name and address of hospital.....	Yes	Yes	No	Yes	Yes
Operations.....	Yes	Yes	No	No	Yes
Dates of.....	No	No	No	No	No
Nature of.....	Yes	No	No	No	Yes
Dental services.....	Yes	Yes	No	Yes	No
Dates of commencement..	No	No	No	Yes	No
Days of treatment during survey period.....	Yes	No	Yes	Yes	No
Medicine.....	No	Yes	No	Yes	No
Home nursing.....	Yes	No	No	Yes	Yes
Still under medical care for each chronic condition.....	No	No	No	No	Yes
Examination for glasses; physiotherapy; osteotherapy; chiropractic and chiropodist treatments.....	Yes	No	No	No	No
Type of condition.....	Yes	No	Yes	Yes	Yes
Continuing.....	Yes	No	Yes	Yes	Yes
New.....	Yes	No	Yes	Yes	Yes
Recurring.....	Yes	No	Yes	No	No
Special questions about alcoholism.....	No	No	No	No	No
Special questions about death.....	No	No	No	Yes	No
Special questions about mental disorders.....	No	No	No	Yes	No
Special questions about pregnancies.....	No	No	No	Yes d/	No
Use of appliances.....	No	No	Yes	No	Yes
Use of habit-forming drugs.	No	Sleeping draughts	No	No	No
Use of tobacco.....	No	Yes	No	No	No
History of vaccinations and immunizations.....	No	Yes	No	No	No
X-rays, volume of exposure to	No	No	No	No	No

a/ Question framed to determine whether condition is present in period of interview, and whether it started more or less than one year ago.

b/ Only whether condition was first noticed in past two weeks, in past threee months, or in past 12 months.

c/ Only when necessary for International Statistical Classification (ISC) coding.

d/ Only when complications require special medical treatment.

Table 3. Methods used for verification and evaluation of data in questionnaires

	Canada	Denmark	England and Wales	Japan	USA
Questionnaires edited for consistency and completeness.....	Yes	Yes	Yes	Yes	Yes
Missing or defective data queried?	Yes	No	Yes	Yes	No
Routine evaluation procedures					
Completness of reporting measured by comparison with records from independent sources?.....	No	Yes; compared with total population	No	No	Yes a/
Re-interview made to measure response errors?.....	No	No	No	No	Yes
Verifications of reported morbidity conditions?.....	Yes	Yes	No	No	No a/
By subsequent medical examination.....	No	No	No	No	No a/
By comparison with doctor's records.....	Yes	Yes	No	No	No a/
By comparison with hospital records.....	No	No	No	No	No a/
Scrutiny of completed questionnaires.....	Yes	Yes	Yes	Yes	Yes
Quality control of coding, punching, and tabulating.....	Yes	Yes	Yes	Yes	Yes
Components of variance measured at each sampling stage?	No	Yes	Yes	No	Yes

Table 3. (cont'd)

	Canada	Denmark	England and Wales	Japan	USA
Comparison of survey statistics made with other independent statistics?.....	With communicable disease statistics, illness in the civil service, hospital statistics, international sources, etc.	With cause of death data; also with morbidity statistics from other sources (tuberculosis, diabetes, ulcer)	All demographic material with existing records	With national census also with Patient Survey (receiving medical services in hospitals and clinics)	With other data at report-writing stage, but survey data not adjusted except that basic ratio estimating procedure makes data consistent with independent controls with respect to age, sex, and race distribution of population

a/ Several special studies based on checks with hospital and physicians' records are being made, but these are not routine verification procedures.

Table 4. Periods covered by questionnaire used in field operations

	Canada	Denmark	England and Wales	Japan	USA
Recall periods for:					
Illnesses.....	Period between interviews, which were usually monthly	1 month prior to interview	Acute 2 months prior to month of interview. Chronic 1 year prior to month of interview	Day of interview for first visit period between subsequent visits a/	Acute: 2 weeks prior to week of interview. Chronic: 1 year prior to week of interview
Injuries.....	Recall period for most variable items was period between interviews which were usually monthly. However, instructions were given that as much information for Household Record Card as possible be recorded on initial visit, with delay until interviewer felt he had confidence of family before asking certain questions		2 months prior to month of interview	When medical attention required, day of interview for first visit period between subsequent visits a/	2 weeks prior to weeks prior to
Residual effects of previous injury or illness.....		not applicable			
Impairments.....					Day of interview
Accidents resulting in injury.....				Not investigated	Accidents 2 weeks prior to week of interview. Injuries causing present residual effects or impairments: any time in past

Table 4. (cont'd)

	Canada	Denmark	England and Wales	Japan	USA
Disability.....				Day of interview for first visit period between subsequent visits a/	Current: 2 weeks prior to week of interview Long-term 1 year for bed days from chronic condition; day of interview for chronic limitation of activity
Health services					
Medical.....		1 Month prior to interview			2 weeks prior to week of interview;
Dental.....					2 weeks prior to week of interview; also visits in past year
Hospital.....					1 year prior to interview
Home nursing.....		Not applicable		Not investigated this way	Day of interview
Economic activity.....		1 month prior to interview		Day of interview	1 year prior to interview
Industry and occupation.....					Not investigated
Income.....		Not applicable		Not investigated	1 year prior to interview
Medical and hospital insurance.....		1 month prior to interview	Not stated	Day of first interview	July 1957-June 1958 for hospitalized persons only: 1-year period

Table 4 (cont'd)

	Canada	Denmark	England and Wales	Japan	USA
Veteran status.....		Not applicable	Not stated	Not investigated	Day of interview
Death		Not applicable	Not investigated	Day of interview for first visit period between subsequent visits a/ For pregnancy where abnormal complications required medical attention	Not investigated
Use of appliances.....				Not investigated	Day of interview
Does period shift forward with delay in interviewing?	No	No	No	No	No
Range of period between end of recall period and date of interview.....	Only recall period between monthly interviews (memory aided by diary)	1-10 days	up to 15 days	Not stated	Median, 2 days; usually completed 1-6ddays; delay as much as 6 days

a/ Usually no more than a day, at most a week.

EDUCATION AND LITERACY

1. In the field of education, household surveys have to date been used principally for collection of statistical data in three areas literacy, educational attainment and school attendance. The following discussion of issues in survey content, design and operations in relation to education will therefore consentrate on issues related to these specific areas. However, it might be useful at the outset to present a very brief overview of the whole field of educational statistics, as background and to understand the role and potentialities of household surveys in this general context.

2. In principle, statistics related to education cover all aspects of the educational system, organization, people involved, activities, results, economic and social features, etc., for which quantitative data are considered necessary, or at least desirable, in a given situation. Two major sectors should be distinguished in this field of statistics, namely:

(a) Statistics related to the educational system per se

(b) Statistics related to educational characteristics of the population.

3. The first sector is very broad, dealing with all types of statistics regarding schools, pupils, teachers, educational finance, etc., and data are normally collected directly from the institutions concerned by means of specific questionnaires or from current administrative records. Most of these data are collected on a yearly basis, but some statistics of a more detailed or specific nature may be obtained through intermittent investigations at regular intervals, say three or five years, or through ad hoc surveys devoted to particular problem areas in the field of education. Historically, these statistics, which we may designate as "traditional" or "institutionalized" educational statistics, have been developed within ministries of education or other national agencies responsible for education to meet the current needs for such data for administrative and planning purposes. Within the overall system of national statistics, the development of this particular field must be characterized as rather slow - especially if compared with demographic and economic statistics. In most countries, developed as well as developing, it is only over the last 20 years or so that educational statistics have been significantly improved and reinforced. Inasmuch as the important role of education for the general economic and social development process has only lately become widely recognized, more emphasis has been put on educational planning and costing, and, by implication, on the required statistical data base therefor.

4. The second major sector of educational statistics is closely related to and in fact overlapping with demographic statistics, as it deals with the impact of the educational system, past and present, on the population in its totality or with regard to selected groups. Such statistical data are consequently in the first instance collected by

means of population censuses and concern primarily literacy, educational attainment and school attendance. In most countries, censuses are taken only at 10-year intervals. However, household surveys on a sample basis are now being more and more used to supplement the infrequent census data with frequent, more up-to-date, and sometimes also more detailed information. This holds true in many fields but what is of interest in this particular field is the fact that such surveys provide an excellent tool for obtaining more timely, reliable and differentiated statistics on educational characteristics of the population. As such data are essential for formulating development plans, not only for education proper but also for economic and social progress in general, where frequent references are made to the "stock of skilled manpower" or "human resources", it is to be hoped that developing countries especially will make more frequent use of household surveys for this particular purpose. The major aim of this chapter is to provide technical guidance and explanations in order to facilitate such undertakings.

5. As for the place of this particular type of educational statistics in the national statistical system, it should be noted that central statistical services, or special census or survey agencies, are responsible for the operation of population censuses and household surveys, even those which include educational data. In this connexion it is appropriate here to make reference to some pertinent features of statistical organization as regards educational statistics.

6. First, in countries with a federal system of government, educational statistics, like all other branches of public statistics, must originally be collected for each state separately and then co-ordinated and compiled at the federal government level. This may, in certain countries with important diversities and inequalities between the educational systems of the different states, be a quite difficult and complex operation.

7. Secondly, in certain countries with a highly centralized statistical organization, the "traditional" sector of educational statistics referred to above, is no longer dealt with by the ministries of education, but by the central statistical services. However, in most developing countries, the responsibility for these "institutionalized" educational statistics still remains with the ministries of education. When household surveys with items on education are organized in such countries, it is therefore particularly important to ensure the closet possible co-ordination and collaboration between the agency in charge of the survey and the ministry of education, which can provide valuable specialized expertise from both statistical and subject-matter viewpoints for the education part of the survey.

A. Scope, purpose, uses and needs for data

1. Scope

(a) Educational characteristics of the total population and of specific population groups

8. It goes without saying that the household survey approach is particularly suitable for collection of educational statistics that illustrate the situation of an individual member of a household with regard to education. Therefore we will discuss here what was called above the second major sector of educational statistics, namely educational characteristics of the population. "Individualized" data on education which can usefully be included in household surveys relate primarily to the following subjects:

Literacy/illiteracy;
Educational attainment;
School attendance;
Access to education.

9. Explanations concerning concepts, definitions, classifications and tabulations are given below in the pertinent sections of this chapter. It should be recalled here that the above types of data may also be collected in population censuses, but household surveys have important advantages. Because of the restricted size of the operation, they can study the subjects in question in more depth, using specific tests where appropriate, so they constitute a very useful complement to large-scale censuses for checking their results and extending the data base.

(b) Statistics related to the educational system

10. As has been explained above, data pertaining to the first major sector of educational statistics are normally collected within the institutional framework in the field of education. It is nevertheless appropriate here to make some brief supplementary reference to these traditional educational statistics in order to place the types of data being collected through household surveys in the right perspective. Institutionalized statistics on education generated by the educational system itself normally relate to the following subjects:

Schools and other institutions (by level and type of education);

Enrolment (by level, type and grade), completion of study course, repetition, drop-out, etc.;

Teaching staff and other personnel;

Educational facilities (buildings, classrooms, laboratories, libraries, equipment, school meal and health services, etc.);

Educational expenditure and finance.

11. It is evident that on the whole the institutionalized data collection method traditionally used for these types of educational statistics is the most appropriate one. Nevertheless household surveys may in a given situation provide an interesting complementary source of information. For example, data on private expenditure on education could be obtained from such surveys to supplement the institutionalized statistics, as these may often be rather incomplete in this respect. Furthermore, on some occasions data regarding the socio-economic situation of pupils might usefully be collected by means of household surveys. The details must be worked out in each particular case, depending on the purpose and the capability of the survey and on specific conditions of the country.

2. Purpose

12. First of all, the purpose of including one or another item on education in a household survey depends on the overall nature and objective of the survey.

(a) Education as main topic of a survey

13. If education is the main topic, or one of several major topics, of a survey, the purpose, in principle, is to provide statistical information concerning the status and development of education (regular, adult and special, of all levels, types and grades) which is considered necessary and which is not elsewhere available. The information should primarily refer to the country as a whole but to the extent possible, depending upon survey design, it is desirable to provide for breakdowns by major regions. Given the scope of the data (as presented in section 1 above) which deal essentially with educational characteristics of the population, the aim of the household survey in this connexion is generally to throw light upon the outcome of the educational process in the country. This is reflected in the educational situation of individuals, that is, acquired knowledge and skills or lack of them.

14. Data concerning the main educational characteristics, literacy and educational attainment, are key indicators to denote progress or critical shortcomings of the educational system over a long period of time. They thus have an important role to play in national plans for the advancement of both regular and adult education. In this connexion it should be noted that literacy and educational attainment statistics may be particularly helpful if appropriate breakdowns by age and by specific categories within the population are provided, such as labour force, ethnic and religious groups, urban-rural, migrant workers, etc.

15. To summarize, for household surveys mainly focussing upon education, the major motivation is to provide data which, along with other types of information, can serve as a basis for administration, planning, evaluation, study and research in the field of education.

(b) Education as ancillary subject of a survey

16. If, on the other hand, education is only an ancillary subject in a household survey, the situation is completely different. The purpose of the data on education is then to provide the means for studying possible interrelationships between the main topic of the survey and, for example, the literacy status and educational level of the individual concerned. Such connections may be anticipated between education and several socio-economic and other characteristics of the household, for example conditions of work and employment, income, housing, nutrition, health, mortality, fertility, etc.

3. Uses

(a) Education as main topic of a survey

17. Beginning again with the first alternative, where education is the main topic or one of the major topics of a household survey the data are generally used for compilations, analyses and special studies devoted to the educational system in the country, including past trends, present situation and prospects for the future. In particular, such basic data and derived statistics can be utilized for evaluation, planning, projections, etc. Literacy data are essential for developing an adequate network of adult education centres to combat illiteracy. Statistics on the educational attainment of the population, total as well as economically active, provide bench-mark data for the process of planning the future educational system to better meet the requirements of overall economic and social development.

(b) Education as ancillary subject of a survey

18. As for the second alternative, with education as an ancillary subject in a household survey, the data on education are primarily used to complement the statistics pertaining to the main topic so as to permit analyses and special studies with a view to throwing light on the interrelationships among educational characteristics of the household and its members on the one hand, and the relevant demographic, economic, social and cultural subjects, etc., on the other. In turn, the results of such analyses and studies can be exploited for planning purposes in various fields which bear upon household living conditions in the country. Again, this means that within the broader perspective of national planning for economic and social development, due consideration will be given to the important role of education for the fulfilment of development plans.

(c) Reliability checks of census data

19. In addition, the data on literacy and educational attainment obtained by means of household surveys can be utilized for testing the reliability of the same types of statistics collected in population censuses. In specialized sample surveys more time and resources can be devoted to a fairly detailed investigation of educational characteristics of household members than is the case in a large-scale statistical operation of the census type, covering a broad variety of different items.

(d) Collaboration with users of data

20. Finally, when considering possible uses of statistics collected through household surveys, it is as important for educational data as for other types of data that very close collaboration be established in planning and throughout the whole survey operation with potential users of the data. The specific needs of administrators, planners and research workers can then be appropriately matched and balanced against the technical knowledge of the survey operators, statistical, financial and personnel resources available, cost-efficiency considerations, etc.

21. It is only through careful preparations of this kind that adequate organization and content of the survey can be ensured, including proper selection of items and a reasonable degree of detail, so that optimal use is made of necessarily very limited resources. This holds true both when education is the main topic of a survey and in cases where it is only an ancillary subject. Normally it cannot be expected that the agency responsible for a household survey and the staff primarily engaged in this operation have a specialized knowledge of the field of education. Hence, for those in charge of such a survey it is essential to maintain the closest possible co-operation not only with subject specialists in the ministry of education and other relevant bodies, but also with statisticians working directly in this field.

4. Needs

(a) General considerations

22. To determine exactly which types of data should be collected in order to best meet the objectives of a given statistical operation is invariably a rather intricate matter. Bearing in mind the scope and purpose of the survey in more general terms and the contemplated use of the data to meet particular needs of planners and administrators, one has to examine very carefully which items should be included in the survey and with what degree of detail. It is a well known fact that exaggerated ambitions and expectations, especially on the part of future users of the data, pose a considerable risk of overburdening the survey with a variety of data and details, all of which are not absolutely indispensable in a given situation. As the resources available normally are very restricted, this might easily lead to a low level of precision and quality in the data collection process and thus be detrmintal also to those data which are really essential for the major purpose of the survey.

23. From this, it is clear that the main emphasis of the survey, on education or on another topic, will have a direct effect upon the selection of items, the classifications to be used, etc. Normally broader content and a greater degree of detail will be aimed at when the survey is focussing on education. However, if the major concern is to study other topics, the education part of a survey must necessarily be limited to a few significant types of data, to illustrate the particular impact of education upon the demographic, economic or social phenomena in question.

(b) Nature of the survey

24. Another criterion with regard to data needs is obviously the nature of the particular survey in the general context of other surveys and other relevant statistical operations in the country. If the survey forms part of a continuing and systematic data collection scheme with surveys conducted on a regular basis and with rotating topics, the education element may at a certain phase of the operation be dealt with more comprehensively and in greater detail. On the other hand, in countries which still do not have continuous survey programmes established, it may be necessary to restrict the education content to a bare minimum of essential and highly significant items in order to cover a greater variety of other subject areas.

(c) Co-operation with users

25. Finally, it must be strongly emphasized again that for difficult decisions regarding the types of data and degree of detail to be retained in the field of education for a given household survey, very close co-operation must be established with those who will use the results of the survey. Only after a careful examination of survey content with users and with specialists in the various fields of statistics will it be possible to arrive at a realistic and balanced decision as to the data needs for the survey, choice of items and degree of detail.

B. Conceptual issues

1. General considerations

26. As is the case with all statistical undertakings, household surveys depend first of all upon a rational, systematic and logical approach with a well delineated universe and clear-cut concepts, definitions and classifications to meet their objectives. An important prerequisite for carrying out such a survey successfully is therefore a strict and concise conceptual framework. For those in charge of a household survey in which data on education are to be collected, it is thus essential to be at least reasonably familiar with common terminology used in this field. To this effect, a few conceptual issues of a more general nature will be discussed in this section. Detailed guidelines concerning suitable standard definitions, methods of measurement and classifications for the principal education items will be presented in section C.

27. It goes without saying that strict adherence to established standards and norms, whether national or international, is of the utmost importance for achieving consistency and comparability with regard to the statistics obtained from household surveys. In countries where national statistical standards already exist for the field of education, they must be taken into account when organizing surveys. This is one important reason for maintaining close collaboration with ministries of education and other relevant agencies, as discussed above.

28. However, many countries, particularly among the developing ones, have not as yet established any comprehensive system of definitions and classifications relating to education. The discussion in this and following sections therefore primarily focusses upon international standards, norms and terminology which are widely used and accepted. This should, however, in no way prevent countries from working towards national definitions and classifications which properly reflect characteristics features of their respective educational systems and particular conditions in each country. At the same time, it is highly desirable to make such national standards as compatible as possible with current international systems in the same field.

2. Basic conceptual issues

(a) Concept of education for statistical purposes

29. A general discussion of conceptual issues should first concentrate on how to delimit the field under study, namely education. The first concern is to draw the boundaries of the universe to be covered as neatly and sharply as possible. Quite a number of different words and concepts may be encountered in education terminology, such as education, instruction, learning, training, study courses and programmes of different levels and types intended for a variety of groups and specific categories within the population and with a broad range of objectives. This is not the place to enter upon a full-fledged discussion of such concepts and definitions nor to attempt a comprehensive definition of the concept education for overall purposes, still less to try to impose a standardized conception of the philosophy aims, or content of education or reflect on its ideological or cultural aspects per se. What is essential in the present connection is to have a simple, deliberately short definition of education, even if it is arbitrary, to meet the practical requirements of statisticians. Such a definition is given in the International Standard Classification of Education (ISCED) established by UNESCO, namely:

> Education is taken to comprise organized and sustained communication designed to bring about learning. (1.36 p.2)

30. For a proper interpretation of this definition, the following supplementary clarifications are provided:

- Communication requires a relationship between two or more persons involving the transfer of information;
- Organized is intended to mean planned in a pattern or sequence with established aims or curricula. It involves an educational agency which organizes the learning situation and/or teachers who are employed (including unpaid volunteers) to consciously organize the communication;
- Sustained is intended to mean that the learning experience has the elements of duration and continuity;

- Learning is taken as any change in behaviour, information, knowledge, understanding, attitudes, skills or capabilities which can be retained and cannot be ascribed to physical growth or to the development of inherited behaviour patterns. (1.36, pp.2-3)

31. Included, therefore, are activities that in some countries and in some languages may not usually be described as education, but rather as training or as cultural development. Excluded, however, are types of communication that are not designed to bring about learning or that are not planned in a pattern or sequence with established aims. Thus, all education involves learning but many forms of learning are not regarded as education. Leisure activities such as recreation, sports, and tourism which are not designed to bring about learning and which do not involve an organised educational agency are excluded from education. "Self-directed learning", "family and socially-directed learning" and "random learning" are excluded because they involve no organized agency or teacher in the above sense. Also excluded are isolated events involving no sustained educational activity, such as one or two public lectures, conferences or meetings entertainment; information, advertising and sales programmes; other social and corporate activities, such as meetings of clubs or associations, and work camps.

(b) Principal types of education

32. From the administrative and organizational point of view, there are three principal types of education to be distinguished:

Regular education;
Adult education;
Special education.

33. The first type of education corresponds to the traditional educational system, at all levels. This is regular school and university education, that is, the system that provides a "ladder" for children and young people who may progress from pre-primary or primary school through university (although many drop out on the way). Education of this type is covered regardless of ownership or sponsorship of the institution concerned (for example, public or private).

34. The second principal category, adult education, comprises education provided to people who are not in the regular school and university system, including adult literacy programmes as well as organized and sustained training for such groups as farmers, craftsmen, industrial workers or homemakers. Programmes of this kind may be arranged by governments, departments, industry, trade unions, the armed forces, and many other agencies. From a statistical point of view, this implies that this category is a highly differentiated and complex one, giving rise to numerous problems with regard to data collection, as the administrative framework is much less rigid and well-defined than for the regular educational system. However, the statistician must face these difficulties and find practicable solutions so that the data collected reflect the factual situation as faithfully as possible.

35. In the past, the adult education sector was in many countries of relatively minor importance from a quantitative point of view. Lately, however, this situation has changed drastically, in line with modern developments in educational policy and planning. It is now widely recognized that education is not an instructional "package" provided to children and young persons in schools and universities, but is a continuing process providing opportunities for enlarging experience, understanding and skill throughout the life of the individual, through "permanent education" or "life-long education".

36. The third category is, in effect, one which overlaps the former two categories in the sense that special education, organized specifically for various groups of handicapped or otherwise exceptional children, adolescents, or adults, may be discerned within both the regular education system and the adult education one.

(c) Content of education

37. It should also be noted that teaching and learning may take place in all kinds of places, for example, in classrooms, community centres, or the open air, and by many methods, for example, lectures, discussions practice work, correspondence, through any appropriate media, for example, books, teaching machines, radio, films, and television. It is thus the content and objectives of a learning process rather than varying organizational and technical arrangements that determine whether any given process should be considered as education or not. In the same way, education in principle embraces all types of pupils and students of all age groups, whether studying on a full-time or on a part-time basis, with varying characteristics and under varying circumstances.

38. As the concept of education is thus primarily focussed upon the content or subject matter of study, it is appropriate here to give a brief presentation of current terminology in that respect.1/

39. The most detailed unit of education is denoted as a course and represents a planned series of learning experiences in a particular range of subject matter or skills offered by a sponsoring agency and undertaken by one or more students.

40. A selection of one or more courses is called a programme, mainly oriented towards a specific field and with an expressed or implied aim, such as qualification for more advanced study or for one or several occupations, or solely an increase in knowledge and understanding. Programmes which are related in terms of level and major subject-matter content may be combined into programme groups, identified in the ISCED system by a five digit code. Programme groups related to the same general subject matter or area are further aggregated into fields (ISCED three digit code).

1/ For more detailed information, reference is made to the ISCED system (1.35, 1.36).

(d) Level of education

41. Finally, there is the concept of level of education (ISCED one digit code), which represents a broad section of the education "ladder", that is, the progression from very elementary to more complicated learning experience, embracing all fields and programme groups that may occur at that particular stage of the progression.

(e) Use of ISCED in household surveys

42. As has been explained, the International Standard Classification of Education (ISCED) is a three-stage classification with a coding system of five digits, providing successive subdivisions from level to field to programme group. This system, adapted as necessary to particular conditions and requirements in the country concerned, may also be used in household surveys for classification of education. The degree of detail will depend on the purpose and nature of the survey. At this stage, it will suffice to indicate some basic elements in the ISCED system.

43. The "education ladder" is made up of the following major levels:

(i) Education at the first level (ISCED category 1), which usually begins between ages 5 and 7 and lasts for about 5 or 6 years;

(ii) Education at the second level, first stage (ISCED category 2), which begins at about age 10-12 and lasts for about 3 years;

(iii) Education at the second level, second stage (ISCED category 3), which begins at about age 13-15 and lasts for about 3 years;

(iv) Education at the third level (ISCED categories 5, 6 and 7), which begins at about age 17-19 and is of varying durations depending upon the stage, as discussed below.

44. A more comprehensive picture of the ISCED level-category pattern is obtained by using the one digit classification in full, as follows:

0 Education preceding the first level
1 Education at the first level
2 Education at the second level, first stage
3 Education at the second level, second stage
5 Education at the third level, first stage, of the type that leads to an award not equivalent to a first university degree
6 Education at the third level, first stage, of the type that leads to a first university degree or equivalent
7 Education at the third level, second stage, of the type that leads to a postgraduate university degree or equivalent
9 Education not definable by level.

45. Here two new categories have been introduced: 0 (for pre-school education, nursery schools, kindergartens or similar institutions) and 9 (for types of education which do not lend themselves to the level concept). In addition, education at the third level has been broken down into three categories to provide for a more differentiated presentation of what has usually been called "university and other higher education", or "post-secondary education".

46. The next step in the ISCED system is the field. A detailed description of that part of the structure is beyond the scope of the present discussion. For this purpose ISCED itself should be consulted (1.35, 1.36). However, by way of example, it may be mentioned that the two second level categories (ISCED codes 2 and 3) can be subdivided into the following major groups of fields:

(i) General education, which stresses the theoretical, philosophic or mathematical aspects of the subjects covered, with little emphasis on technical subjects dealing with practical skills (ISCED codes 201 and 301);

(ii) Teacher training (ISCED codes 214 and 314);

(iii) Other education at the second level (ISCED codes 224, 234, 250, 252, 262, 266, 270, 278, 289, 318, 326, 334, 350, 352, 354, 362, 366, 370, 378 and 389).

47. With a similar subdivision at the third level, one arrives at an intermediate grouping of levels and fields comprising 18 categories (1.36, pp. 31-32), which might be of interest for some household surveys. The full three digit classification contains some 100 categories and is thus relevant mainly for specialized surveys concentrating on an in-depth study of education.

3. Other conceptual issues of a general nature

48. After having dealt above with certain central conceptual issues in the field of education, this section will refer to a few concepts and terminological usages typical for educational statistics as such. Some familiarity with such concepts and expressions is helpful to place the whole area of educational statistics in perspective.

(a) Stock/flow

49. In the past, educational statistics were predominantly of a static character. Figures were presented concerning the number of pupils registered (enrolment) and teachers engaged at given points in time, without any attempt to describe in more detail the continuous process of evolution and change in these respects during and, above all, between the school years.

50. In modern educational statistics, there is a very clear trend towards a more dynamic approach. To this end, the concepts of stock and flow have been introduced. Instead of simply observing and reporting

net changes in enrolment figures (stock data) from one school year to the following one, educational statisticians have recently developed and improved methods to follow cohorts of pupils year by year as they progress through the educational system (flow data). These are used to describe what happens to cohorts at each particular step of their "school career". Similarly, the flow approach can be used for teachers, analysing the movements into, within and out of the teaching staff.

51. Even with modest means, it is feasible to arrive at a fairly satisfactory data base for flow studies. With a view to assisting countries where only stock data still exist, UNESCO has developed a methodology for flow analyses giving special attention to evaluation of "educational wastage". Another methodology of interest to developing countries has been designed using specific registration numbers for pupils, identifying them with the original cohort of school entry (1.38). This coding system represents a substitute for a complete individualized data system and is intended for countries where lack of computer facilities makes the latter, more sophisticated approach impractical.

(b) Educational wastage

52. The important concept 'educational wastage' is now currently used as a kind of "main heading" for the two phenomena "repetition" and "drop-out". For statistical purposes, the following definitions may be used (2.65, p.16):

- Drop-out or school desertion: leaving school before the completion of a given stage of education or leaving at some intermediate or non-terminal point in a cycle of schooling,
- Repetition: a year spent by a pupil doing the same work in the same grade as in his previous year in school;
- Educational wastage: incidence, in a country's education system, of drop-out and repetition.

53. In some countries very high levels of drop-out and repetition constitute an extremely serious problem. Continuous analyses are then indispensable to evaluate what might be called the internal efficiency or the "output' of the educational system. This is all the more important in situations where available resources are severely limited and where it therefore is imperative to use them with maximum efficiency.

(c) Educational disparities

54. Another conceptual issue which also focusses on potential deficiencies of the educational system concerns educational disparities. Here it is the social aspect that is decisive, namely questions of equal access to various kinds of education, the position of different population groups vis-à-vis the educational infrastructure, handicaps of a geographical, economic, cultural, ethnical nature, etc. This is an area which so far has been much neglected, as traditional educational

statistics have tended to study the education system in terms of structures and institutions. This is an information gap which can be filled by using the household survey approach, as the unit of observation should be the individual. The concern is with the individual's relation to the supply of educational services that is, how people avail or do not avail themselves of educational facilities. In this connexion, concepts such as access to education, distance to school, etc., also come to the fore.

55. This is a relatively new and unexplored sector of educational statistics which still lacks a generally accepted conceptual framework and a standardized terminology. However, this situation is changing as interested survey organizers undertake new work in this area. Various approaches are being tried in order to shed light on a complex problem of great importance for educational policy and planning.

C. Principal items of information

1. Items especially suitable for household surveys

56. As has already been pointed out above, household surveys lend themselves particularly well to collecting data on educational characteristics of the population and similar types of statistics.

57. Among items to be included in such surveys the following are of first importance:

- literacy
- educational attainment
- school attendance.

58. Depending upon the specific nature and purpose of the survey and upon particular conditions and requirements, other items might also be covered, such as educational qualifications, access to education (distance to schools and other educational institutions), etc. It must be recalled that it is not possible to present any firm recommendations on survey content with respect to the vast fields of education because of the highly differentiated and flexible character of household surveys and the great variety of specific national circumstances throughout the world.

2. Items normally included in educational statistics collected directly from educational institutions

59. In section B.1, it was noted that institutionalized educational statistics as a rule will continue to be collected in the traditional way. A brief list of the most important types of data for this category was given in that section as background information, in order to convey some idea of the whole universe of educational statistics. It has furthermore been indicated that in particular cases, for example concerning private expenditure on education or the socio-economic characteristics of pupils and their households, the household survey

approach might also be halpful in collecting data pertaining to the "traditional" sector of educational statistics.

60. The next section presents a comprehensive review of standards recommended for the typical household survey items on education. The following section will refer briefly to similar standards for a few key elements in traditional educational statistics.

3. International standards relating to educational characteristics of the population

61. The following standard definitions, methods of measurement, and classifications are based on current international recommendations or, if recommendations do not exist, on commonly accepted practice. In cases where more detailed and specific information might be required for elaborating survey programmes, the reader is referred directly to the sources quoted.

(a) Literacy/illiteracy (1.42)

(i) Definitions

62. A person is literate who can with understanding both read and write a short simple statement on his everyday life.

63. A person is illiterate who cannot with understanding both read and write a short simple statement on his everyday life.

64. A person is functionally literate who can engage in all those activities in which literacy is required for effective functioning of his group and community and also for enabling him to continue to use reading, writing and calculation for his own and the community s development.

65. A person is functionally illiterate who cannot engage in all those activities in which literacy is required for effective functioning of his group and community and also for enabling him to continue to use reading, writing and calculation for his own and the community's development.

(ii) Methods of measurement

66. To determine the number of literates (or functional literates) and illiterates (or functional illiterates), any of the following methods could be used:

(a) Ask a question or questions pertinent to the definitions given above, in a complete census or sample survey of the population

(b) Use a standardized test of literacy (or functional literacy) in a special survey. This method could be used to verify data obtained by other means or to correct bias in other returns;

(c) When none of the above is possible prepare estimates based on:

(i) Special censuses or sample surveys on the extent of school enrolment;

(ii) Regular school statistics in relation to demographic data;

(iii) Data on educational attainment of the population.

(iii) Classification

67. The population aged 10 years and over should be classified first into two groups: literates and illiterates. Where appropriate, functional illiterates should also be distinguished.

68. Each of these groups should be classified by sex, and also by age in the following groups: 10-14, 15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65 years and over.

69. Additional classifications should be made, where appropriate, for:

(a) Urban and rural populations;

(b) Such ethnic groups as are usually distinguished within a state for statistical purposes;

(c) Social groups.

(b) Educational attainment of the population (1.42, pp.2-3)

(i) Definition

70. The educational attainment of a person is the highest grade completed and/or the highest level of education attained or completed by the person in the system of regular, special and adult education of his own or some other State.

(ii) Methods of measurement

71. To measure the educational attainment of the population, the following methods could be used:

(a) Ask a question or questions pertinent to the definition given above, at a complete census or sample survey of the population;

(b) Where this is impossible, prepare estimates based on:

(i) Data from previous censuses or surveys;

(ii) Records over a number of years of school enrolment, of examination, of school-leaving certificates, and of degrees and diplomas granted.

(iii) Classification

72. The population 15 years old and over should first be classified by educational attainment, expressed preferably in terms of highest grade completed, but at least in terms of level of education attained or completed. Wherever possible, distinction should also be made among different fields of study at each level.

73. Each of these groups should be classified by sex and also by age in the following groups: 15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65 years and over.

74. Additional classification should be prepared, where appropriate, for:

(a) Urban and rural populations;

(b) Such ethnic groups as are usually distinguished within a state for statistical purposes;

(c) Social groups.

(c) School attendance (1.42)

(i) Definition

75. School attendance is defined as attendance at any accredited educational institution, public or private, for systematic instruction at any level of education at the time of the survey, or, if the survey is taken during the school vacation period, at the end of the last school year. Instruction in particular skills which is not part of the recognized educational structure of the country (for example, in-service training courses in factories), is not considered as school attendance for purposes of a survey.

(ii) Methods of measurement

76. Ask a question or questions pertinent to the definition given above, in the first place to all persons between the usual age for entrance into the first level of education and 24 years of age inclusive. If it is considered likely that a significant number of older persons is attending school, such information may also be collected for persons over 24 years of age. This is particularly the case in countries which may also wish to investigate attendance at facilities providing adult education, outside the regular school and university system.

77. It should be borne in mind that there is a very great variety of durations in adult education programmes. Accordingly, a question should be asked on the duration in hours of programmes for adults. The population in the relevant age range, as indicated above, should first be classified by level of education attended. Whenever possible, a classification by grade and broad fields of study should also be made. If adult education has been included, this sector of education should

be shown separately and, preferably, should be classified according to ISCED 5 digit (or 3 digit) codes. In addition, those attending school should be classified by sex and 5 year age groups. Additional classifications may be introduced, where appropriate, for urban and rural populations, ethnic and social groups, etc.

(d) Educational qualifications (1.42)

(i) Definition

78. Educational qualifications are the qualifications (that is, degrees, diplomas, certificates, etc.) that an individual has acquired, whether by full-time study, part-time study or private study whether conferred in the home country or abroad; and whether conferred by educational authorities, special examining bodies or professional bodies. The acquisition of an educational qualification implies the successful completion of a course of study.

(ii) Method of measurement

79. Ask a question or questions pertinent to the definition given above, primarily to all persons having completed a course of study at the third level of education (ISCED codes 5, 6 and 7) regardless of the age of the person.

4. Other international standards in the field of education

80. As has already been emphasized, data on education collected through household surveys should not be seen in isolation but rather as an integral part of an overall statistical information system for the field of education as a whole. The international standards most directly relevant to educational data which might be collected in household surveys were discussed above. Standard definitions relating to other key elements in education statistics are summarized below (1.42, pp. 3-5):

(a) A pupil (student) is a person enrolled and/or registered in a programme of education;

(b) A teacher is anyone employed, even part-time and/or without remuneration, to communicate knowledge, skills, etc.:

(i) A full-time teacher is a person engaged in teaching for a number of hours customarily regarded as full-time at the particular level of education in each state;

(ii) A part-time teacher is one who is not a full-time teacher;

(c) A grade is a stage of instruction usually covered in the course of a school year;

(d) A class is a group of pupils (students) who are usually instructed together by a teacher or by several teachers;

(e) A school (education institutions) is a group of pupils (students) or one or more grades organized to receive instruction of a given type and level under one teacher, or of various types and/or levels under more than one teacher, under the direct supervision of the head of the establishment (the school or educational institution is often the unit from which statistics may be secured):

(i) A public school is operated by a public authority (national, federal, state or provincial, or local) whatever the origin of its financial resources;

(ii) A private school is a school not operated by a public authority whether or not it receives support from such authorities. Private schools may be defined as aided or non-aided, respectively, according as they derive or do not derive financial support from public authorities;

(f) The compulsory school-age population is the total population between the age limits of compulsory full-time education;

(g) Educational finance:

(i) Receipts refer to cash received by or made available to or for schools, including appropriations, subventions, fees, cash value of property received as gifts, etc.;

(ii) Expenditures refer to financial charges incurred by or on behalf of schools for goods and services. Expenditures, in turn, may be further subdivided into:

Current expenditures;
Capital expenditures;
Loan transactions.

81. Again, it should be recalled that the major instrument for classification regarding these various elements is the International Standard Classification of Education (ISCED) (1.35, 1.36), the main features of which have already been dealt with above.

D. Special sampling, measurement and operational issues

1. Sampling issues

82. Regarding sample design and related preparations for household surveys in general, background information and technical advice is provided in part one of the Handbook. For surveys which deal with education, it has to be examined in each particular case what further considerations may be required with due regard to the specific purpose and character of the survey, the role and importance of education and other subjects included in it, regional and local conditions, etc.

83. In the case of household surveys which include education merely as an ancillary subject, such considerations will naturally depend first and foremost on the topic of major concern for the survey. Nevertheless, proper attention must at the same time be given to possible side-effects that the chosen sample design might have with respect to the educational characteristics of the population groups under study.

84. When education constitutes the main topic of a household survey or survey module or round, it is evident that decision concerning sample design primarily have to be made after careful consideration of their practical implications for the various educational features to be investigated. As educational infrastructures and related circumstances vary very much from country to country, it is, however, difficult to establish any firm and universally applicable rules and guidelines as to how to proceed in order to ensure adequate representativeness of the sample in this respect. Moreover, the household survey approach has so far not been much used in the field of education, and practical experience is thus rather limited. It is therfore necessary to tackle this problem from case to case.

85. The needs for specific compilations and analyses are also of importance for the sampling arrangements to be made. If for example, the survey aims at studying the educational situation of relatively small population groups such as ethnic and religious minorities, migrant workers, etc., particular attention has to be given to the adequacy of the sample.

86. A particular sampling issue concerns literacy and related problems of data quality. In order to ensure reasonably reliable information on literacy, enumerators should use some simple test material to check, where required, the ability of each household member aged 10 and over to read and write. In certain cases, it may even be considered necessary to have recourse to more sophisticated test procedures as a complementary control measure. Such an in-depth investigation of literacy status implies, however, considerable demands on time and specialized personnel and could normally not be extended to more than a limited number of respondents. At the first stage of drawing up the sample design for the survey, the needs for this kind of reliablity checking should be considered and provision made, where required, for an appropriately chosen subsample for this particular purpose.

2. Measurement issues

87. International recommendations concerning methods of measurement have been presented above, dealing with some of the educational characteristics of the population. Some further explanations and suggestions, especially regarding those concepts which may create specific difficulties for survey organizers, are given below.

(a) Literacy

88. At the outset, it should be emphasized that substantial experience in population censuses and special surveys indicates that the measurement of literacy presents quite particular problems. The concept itself may be given varying interpretations by those planning a survey, the enumerators, heads of households and the persons whose level of literacy is being measured. In effect, literacy in this connexion should be seen as a continuous variable, ranging from the most elementary and mechanical ability to read and write a few words with great pains and no or little understanding to well developed reading and writing skills that can be utilized for the most advanced purposes of communication. It is obvious that a person who can, for example, write only his name and some numerals and/or read just a few very simple words like "eyes", "no", "good" or "bad", is not literate in any meaningful sense. The crucial problem is therefore to determine a threshold level within this wide range, corresponding to a reading and writing ability that really makes some sense for the individual concerned by enabling him to communicate with other people and to be informed by means of written language, at least in a very modest way.

89. The two difinitions given in subsection D.3 (a) above of a literate person and of an illiterate person, are precisely formulated with the intention of describing a "bare minimum" level of literacy. The key words are "with understanding", to eliminate any kind of merely mechanical ability (rote learning) with practical meaning, "short simple statement", to indicate a quite humble level of communication, and "everyday life", to signify that the message refers only to things that are very elementary and familiar to the person concerned.

90. Ever since this first definition of literacy was introduced by UNESCO in 1958, the need for a more qualified concept of literacy has been felt to supplement the above "bare minimum" level. Especially in formulating objectives for mass literacy campaigns at the international as well as at the national levels, it has been necessary to aim at a more ambitious degree of literacy, focusing primarily on the proper "functioning" of the individual in relation to his community and socio-economic situation. Thus the concept of "functional literacy" has been created, corresponding to the third and fourth definitions in Section D above, definitions which were included in the 1978 revision of the UNESCO recommendation (1.42).

91. Functional literacy represents a significantly higher level of literacy which is clearly "work-oriented" and which puts special emphasis on the continuous use of the acquired ability in reading, writing and also calculation for practical purposes. The skills must be sufficiently advanced to make it possible for the individual to participate fully and effectively in activities commonly occurring in his life situation that require a reasonable capability of communicating by written language. This is applicable to both occupational activities and community life. A functionally literate person must, for example, be able to receive written instructions for his work, read newspapers, notices, advertisements, popular literature, etc., write ordinary letters, notes and messages, and make elementary calculations, keep simple accounts, etc.

92. In view of this dualistic approach to the level of literacy, it is especially important that a sharp distinction be consistently maintained between the concepts literacy and functional literacy. The relevant questions in the survey form should closely follow the respective definitions and the enumerators must be provided with comprehensive instructions on how to ensure a correct interpretation of the replies obtained. It is also essential that such questions are not just answered by the head of the household on behalf of its other members but put directly to each individual aged 10 years and over. The enumerator will in this way have a better opportunity of controlling the answers received. The method of private interviews may also have the advantage of overcoming certain psychological inhibitions and prestige attitudes which tend to disguise the real literacy status.

93. In order to measure the literacy level of respondents as accurately as possible, enumerators should take every opportunity of verifying the received replies by means of various control questions. To this effect, it is appropriate that questions on educational attainment precede those on literacy so that immediate comparisons of the replies can be made and apparent mistakes corrected. It is obviously not possible to give any firm rules about the relationship between completed grades of regular schooling and the two literacy concepts in question as educational systems and other circumstances vary greatly from country to country. As a very tentative indication, however, it might be said that for an average developing country the level of literacy would roughly correspond to 3-4 years of primary schooling (repetition not included), and functional literacy to 5-6 years of such schooling. For those having participated in adult literacy courses instead, the corresponding time equivalents might be estimated, very approximately, at some 100 hours and 300 hours, respectively. However, there may be quite considerable individual variations. Even a respectable amount of regular or adult education is not a guarantee in itself against relapse into illiteracy when motivation or appropriate reading and writing material is lacking.

94. Other ways must therefore be sought to check and validate the literacy status claimed by the respondent. It is of great help if oral control questions can be combined with at least some simple test material, for example a set of "literacy cards" to be shown to respondents after they have answered the main literacy question(s). Such cards might contain a short printed paragraph followed by a few related questions to which the respondent should give a brief reply, if possible also in writing, thereby providing concrete proof of his ability to read and write with understanding. With the help of special instructions for the evaluation of these answers, the enumerators would then be able to determine whether the respondent was to be considered literate or not.

95. While some rather simple test material of this kind substantially increases the validity of data on literacy, it is evident that still better measurement of literacy levels can be obtained by introducing standardized achievement tests of a more comprehensive nature. By

standardized test in this connexion is meant a test for which the content has been selected and checked empirically, norms have been established, uniform methods of administering and scoring have been developed, and which may be scored with a relatively high degree of objectivity. Especially when dealing with functional literacy, such an achievement test, covering not only reading and writing but also arithmetic, is an ideal measurement tool. However, more elaborate tests of this kind are quite resource-demanding in terms of specialized personnel, time allocation, etc., and they are usually administered in schools. As has already been indicated in sub-section 1 above, they are therefore as a rule utilized only on a subsample basis. Nevertheless such standardized tests have a very important role to play in checking reliability with regard to the measurement methods ordinarily used for the survey as a whole.

96. Finally, it should be mentioned that the international guidelines on methods of measurement relating to literacy also include methods for the preparation of estimates. Direct information on literacy obtained through population censuses and special surveys is often rather scarce, and it may sometimes be helpful to have recourse to estimates to fill gaps in time series and for establishing projections. For technical advice, reference is made to the relevant UNESCO documents (2.69). When preparing such estimates, specific national conditions should be taken into due account, for example the impact of large-scale literacy campaigns

(b) Educational attainment

97. To measure the educational attainment of a person requires, in principle, some investigation, more or less detailed, of his "school history" to determine which segment or point he has reached on the "educational ladder". This can, in practice be done with different degrees of precision, as is apparent from the definition given in section D above, giving a choice with regard to "the highest grade completed and/or the highest level of education attained or completed". In most cases, especially when education is a major topic in a survey, it is preferable to use the more precise alternative, that is the highest grade completed at the highest level of education attained. Particularly for many developing countries, where drop-out rates are often high, restricting the information to the highest level completed implies extremely crude measurement of the stock of educated people and would, for example, deprive the survey organizers of useful control possibilities with respect to literacy responses. Even among those who have not completed the first level of education as such, there may be a considerable number who have completed at least grade 3 or 4, who are thus potentially literate.

98. Once the choice has been made as to the degree of precision desired, it is important that the questions in the survey form and the related instructions are formulated in a clear and concise way so as not to give rise to any misinterpretation. The key words in the definition of "level" and "grade" have been given above in sections C and D.

Reference was also made there to other possibilities of using the ISCED system, including, for example, groupings by fields of study. This might be of particular interest in connexion with manpower studies, but the degree of detail to be introduced will naturally depend on the principal orientation and purpose of the survey.

99. It is very important to avoid any confusion between "highest grade completed" and number of years spent in school. A person may very well have stayed at school during seven years but, because of repetition of several grades, reached only grade 4 before leaving. For the measurement of educational attainment it is not the amount of time as such spent in educational institutions that is relevant, but instead the years successfully completed in terms of the educational level/grade structure.

100. It is also important to note that educational attainment is not confined to regular education only but should comprise adult and special education as well. Enumerators must devote special attention to thes categories of education, which might otherwise easily be overlooked, and thereby also try to identify to the extent possible the level/grade equivalence in relation to the regular education system.

101. Finally, for those who are found not to have participated in any form of organized education, a special category "no schooling" should be provided for in the reporting form.

(c) School attendance

102. It should be noted that the term "attendance" must not in this context be given too narrow an interpretation, in the sense of physical presence at an educational institution on the day of the interview. What is essential is to identify all persons who in principle are to be considered as pupils or students, that is, enrolled and/or registered in a programme of education. Temporary absence because of illness or for other reasons should be disregarded.

3. Operational issues

103. This section will discuss some practical questions regarding the organization of data collection and field work in connexion with u surveys covering educational topics.

104. First of all, it should be recognized that education is a relatively rare subject in household surveys and that survey organizers, enumerators and respondents as a rule are not sufficiently familiar with educational terminology, concepts and structures. Special care must be taken to ensure a clear, concise and consistent presentation of questions relating to education in the survey form. It is essential to pre-test the questionnaire in order to find out whether these questions are properly understood and whether the answers can be considered as reasonably correct. In particular for countries with no or little experience of household surveys on education, pilot studies are indispensable to make sure that the envisaged survey is practicable in terms of operating

procedures, time, budget and personnel allocation etc. The findings of such studies, together with experience, should be fully utilized for amending and improving questionnaires and instructions, as well as for preparing manuals and organizing training programmes for enumerators so as to achieve a technically correct and meaningful treatment of the education component of the survey. To that effect it is of great importance that the agency in charge of the survey operation maintain very close co-operation with the relevant educational authorities, particularly with the ministry of education, but where required with local educational agencies, school inspectors, etc., as well.

105. One particular form of educational involvement in survey planning and operation is achieved by recruiting enumerators from the teaching profession and/or among students and pupils among students and pupils in higher and secondary education, as appropriate. This type of arrangement has been practised in some countries where conditions are favourable and where in particular the survey operations can be efficiently synchronized with the respective educational activities. Undoubtedly, an educated field staff, often possessing valuable knowledge and experience of local conditions of relevance to the survey, is an asset for the survey project as a whole. In particular, it can be helpful in providing the required expertise for an adequate treatment of the education component.

106. What has been said so far about operational issues concerning educational subjects in general applies with even stronger emphasis with regard to literacy, which as has been discussed above, is a relatively complex subject. Specific interview techniques and tools have to be developed, taking into due account relevant national, cultural and, where required, linguistic factors. The enumerators must be carefully prepared and trained in applying such techniques and methods of measurement in order to determine the degree of literacy of respondents as accurately as possible. In particular, detailed instructions should be given concerning the use of control questions and simple test material, "literacy cards" and the like, for checking the reliability of the answers. In the case of surveys with a dualistic approach, that is, literacy and functional literacy, enumerators must be given specialized training in distinguishing effectively between these two levels of literacy.

107. For the preparation of materials and the organization of training programmes, survey organizers should seek the advice and co-operation of the pertinent educational authorities, literacy boards etc. If standardized tests are to be used for a thorough examination of literacy levels, materials and procedures should be worked out in close collaboration with competent institutions specializing in pedagogics and psychometry.

108. Finally, it may be noted that some multi-lingual countries find it appropriate to introduce in their survey programme a specification of literacy according to language. On such example is given in the National Literacy Survey in Kenya in 1976 2/, in which the principal question concerned the ability to read and write in the mother tongue. In addition, similar questions regarding Swahili and English were to be put to the respondents actually speaking those languages. The experience of this survey shows, however, that the diversified approach to literacy easily can lead to misunderstandings. In cases of this kind, it is therefore essential that the relevant questions are very clear and precise and that enumerators are given explicit instructions to keep the respective linguistic groups separate.

E. Possibilities of tabulation and analysis

1. Tabulations

109. The possibilities for compiling tables on various educational items collected in household surveys depend on quite a number of factors, the principal purpose and nature of the survey, the specific place and role of education in this context and whether it is a major topic or an ancillary subject, the various educational items included, the data base and its degree of detail, the need for and feasibility of cross-classifications in different respects, sampling considerations, etc. Careful attention must evidently also be given to the anticipated use of the tables for analysis and research purposes. A comprehensive and detailed tabulation programme can therefore only be drawn up for a particular survey programme taking into due consideration all these factors as well as specific national circumstances with regard to educational structures and socio-economic conditions.

110. This sub-section therefore provides only a tentative list of a few basic and typical tabulations, corresponding to commonly occur requirements for interpretation and analysis of survey data relating to the various educational items. These might provide a starting point for further adaptation to specific conditions and needs in each programme.

(a) Literacy

111. Initial tabulations might include:

(i) Population 10 years of age and over, by literacy, age group (10-14 years, 15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65 years and over) and sex;

(ii) Population 15 years of age and over, by literacy, sex, urban and rural areas, and possibly major regions;

(iii) Population 15 years of age and over, by literacy, sex and ethnic group

(iv) Population 15 years of age and over, by literacy, sex and social group.

2/ Conducted as part of the National Integrated Sample Survey Programme (2.28,2.29).

Special observations

112. For surveys dealing with functional literacy, similar tabulations may be used for that subject.

113. In some developing countries, the average age for school entrance may be rather high, which means that data on literacy for the lowest age group, 10-14 years, are not very meaningful and can be omitted. The rest of the age classification indicated above is, however, a minimum requirement for purposes of international comparisons, with the age of 15 years as a common lower limit.

14. In multi-lingual countries, specification by "language in which literate" may be introduced in the tabulations.

115. In the case of multi-subject surveys, tables concerning literacy might, in accordance with the purpose and character of such surveys, also include cross-classifications with demographic and socio-economic subjects of special relevance, such as fertility, employment, economic activity, occupation, etc.

(b) Educational attainment

116. Initial tabulations might include:

(i) Population 15 years of age and over, by educational attainment, age group (15-19 years, 20-24, 25-34, 35-44, 45-54, 55-64, 65 years and over) and sex;

(ii) Population 15 years of age and over, by educational attainment, sex, urban and rural areas and possibly major regions;

(iii) Population 15 years of age and over, by educational attainment, sex and ethnic group;

(iv) Population 15 years of age and over, by educational attainment, sex and social group.

Special observations

117. In principle, these tabulations should concern the total population of the particular age group including both persons with completed education (that is, having left the educational system) and children and young people still undergoing education at the time of the survey (that is, remaining within this system). However, if the latter category includes a significant proportion of the population aged 15-24, it is recommended that the age group 25 years of age and over be used in preference to that of 15 years and over as the educational status of those still in school is subject to change in the future.

118. Depending on the purpose and orientation of the survey, the tabulation programme might also provide for cross-classifications between educational attainment and certain demographic and socio-economic subjects of particular interest, such as fertility, mortality, employment, economic activity, income, etc.

(c) Educational qualifications

119. Initial tabulations might include:

(i) Population that has successfully completed a course of study at the third level of education, by educational qualifications (degrees, diplomas, certificates, etc.), age group (15-19 years 20-24, 25-34, 35-44, 45-54, 55-64, 65 years and over) and sex;

(ii) Population that has successfully completed a course of study at the third level of education, by educational qualifications, sex, urban and rural areas and possibly major regions.

Special observations

120. In case the survey also covers qualifications (diplomas and certificates) pertaining to the second stage of second level education, similar tables should be prepared for the persons having acquired such qualifications.

121. With due regard to the purpose and coverage of the survey and to sample error, educational qualifications might be cross-classified with, for example, economic activity, occupation and income.

(d) Distance from school

122. Initial tabulations might include:

(i) Number of households and population in households by distance from the home to the nearest school at the first level of education, by urban and rural areas, and by major regions;

(ii) Number of households and population in households, by distance from the home to the nearest school at the second level of education, by urban and rural areas and possibly by major regions.

Special observations

123. Depending on the local situation, the distance from the home to school may be classified by length of way (for example, 0-1 km, 1-3, 3-6, 6-10, 10 km and over), or , especially where school transport facilities exist, by travel time (for example, 0-10 minutes, 10-30, 30-60, 60-100, 100 minutes and over).

124. These examples of distance classifications are intended for schools at the first level of education. With regard to second level schools, where distances are usually of a quite different order of magnitude, the classification has to be adapted to the actual conditions. Moreover, in countries with a very sparse network of such schools, tabulation (ii) would be of a rather theoretical interest insofar as rural areas are concerned.

125. Where special school travel and boarding facilities are organized, a supplementary table regarding the availability and use of such facilities should be introduced.

(e) General remarks

126. Finally, it must be recalled that the various tabulations briefly outlined above are only meant to serve as illustrative examples. The detailed choice and design of tables for a given survey will necessarily depend on the needs, potentialities and constraints of that particular survey. It is especially important to ensure that the degree of disaggregation in tabulations and cross-classifications is kept within reasonable and realistic limits, taking sampling error into account. Too detailed breakdowns may lead to so many very small figures as to be meaningless.

2. Interpretation and analysis

127. In reviewing possibilities of interpretation and analysis of survey data with regard to educational features and their relationships with other subjects covered by the survey, a few general observations may be kept in mind.

128. Theoretically, different analytical approaches are conceivable, but survey organizers must make their final choices from a purely practical standpoint, taking into consideration the specific purpose and nature of the survey. It is therefore necessary from the first stage of planning to examine carefully which methods of analysis are most appropriate in a given survey situation and to adapt data collection and processing accordingly.

129. A properly designed tabulation programme constitutes in itself a basic instrument for interpretation and analysis of the data. Appropriate breakdowns and cross-classifications in these tables are essential for studying relationships between educational characteristics and certain demographic and socio-economic factors, such as age, sex, occupation, income, etc.

130. On the basis of the material thus available, different statistical measures, such as percentages, rates, ratios, means, medians, standard deviations, etc., can be computed for analyses and studies. Some typical examples of such derived statistics are given below, together with some further suggestions and explanations.

(a) Literacy

131. The two commonly used statistical measures related to literacy are complementary. The positive one, the literacy rate, represents the percentage of a given population which is literate. The negative measure, the illiteracy rate, gives the percentage of the population concerned which is illiterate. These rates have identical meanings, but there is a tendency to prefer the latter as the more "striking" measure, particularly in countries where mass illiteracy constitutes a serious problem.

132. Specific literacy and illiteracy rates may be calculated for adults age 15 and over and any other age range age 10 and over such as 10-15, 15-19, 20-24, 25-34, etc.

133. If functional literacy is covered by the survey, specific rates should be computed for that item in the same way.

134. It should be emphasized that these rates can be used for a variety of analytical purposes, not only within the static context of comparing different population groups for a survey conducted at a given point in time but also in a dynamic sense, studying changes in literacy over time on the basis of data from a series of surveys and censuses. Such time series are particularly important in connexion with educational reforms and massive literacy campaigns in many developing countries, as they provide basic statistics for the evaluation of progress towards the eradication of illiteracy and other educational objectives.

(b) Educational attainment

135. The most direct and simple method of analysis for this item is to use percentage distributions by educational attainment for different population groups, in order to illustrate and compare their educational profiles. In effect, each percentage in such a distribution can be considered as an educational attainment rate for a specific level, or level and grade as the case may be, within the population group concerned.

136. It should be noted that the percentage rate for each category or grade of attainment relates to the completion of that grade, no more and no less. In order to measure the stock of educated people having completed at least a given grade, a cumulative grade attainment rate can be used.

137. Where a full distribution by grade is available, it is also possible to introduce a central tendency measure, for example the median number of school years completed (taking care not to include repeated years). This measure may be defined as the value which divides the distribution of the population by educational attainment (expressed as the number of school years completed and treated as a continuous quantitative variable) into two equal parts, one half of the cases falling below this value and the other half exceeding it.

The median number of school years completed may be quite useful for the purpose of comparisons between different population groups and over time. However, for countries where the majority of the population has no or very little schooling, this measure is not very useful. In the extreme case, it can even assume the value of 0 (more than 50 per cent of the population with no schooling at all).

138. As for orientation of the analytical work, it may be mentioned that in multi-subject surveys there are particularly interesting possibilities for studying relationships between educational attainment and certain demographic and socio-economic factors, including fertility, economic activity, income, etc.

(c) Educational qualifications

139. Methods of presentation and analysis for this item have to be rather flexible, depending on the particular survey, degree of detail of available data, type of classification used, etc.

140. In general, at least some overall educational qualification rates should be calculated to indicate the relative order of magnitude of the stock of persons with special qualifications for relevant broad age groups of the population. To the extent possible further differentiation should be made by fields of specialization in order to allow comparisons with the needs for skilled personnel in various economic and social sectors. However, as has been mentioned above, caution should be exercized not to go into too detailed breakdowns of this usually rather small category of highly educated people, taking the sampling error into proper consideration.

(d) Distance from school

141. A percentage distribution of households as well as population by distance from school is the natural basis for studying the physical aspect of access to education. In addition, the median distance, expressed in physical distance or in travel time, as appropriate, could be used as a central tendency measure so not to overrate extreme distances.

142. In the analysis pertinent information on school transport and boarding facilities should be taken into account. Useful indicators in this respect could be the proportions of households having access to such facilities on the one hand, and of households actually utilizing them on the other.

F. Methods of evaluation of data quality

143. Data on education collected by means of household surveys are inevitably, like other types of data in such surveys, subject to errors of various kinds. Many questions concerning the quality of survey data have already been discussed in part one of the Handbook. In the present section, issues which relate specifically to data quality in the field of education are discussed.

144. As has already been noted, education has to date been relatively rare as a primary subject in household surveys, and as a result experience regarding the quality of educational data collected by means of such surveys is rather limited. There are nevertheless some findings which tend to indicate that data on education are subject to particular risks of error because respondents, and even enumerators and survey organizers often have not been familiar with educational topics and the related, specific terminology. In addition, it is possible that respondents sometimes deliberately give inaccurate answers, for example concerning literacy and educational attainment, for reasons of personal prestige.

145. In these circumstances, it is evident that the problem of data quality has to be given special attention during the whole survey operation. Various methods can be used in order to reduce the risk of errors as much as possible, depending on the particular survey situation, personnel and other resources available, supplementary sources of information for comparison and control, etc.

146. Primarily, it is of great importance that all possible efforts be made to ensure continuous evaluation of data quality throughout field operations, data processing, compilation of tables, preparation of reports and analysis. Comparisons should be made wherever possible with data collected in other ways, for example in population censuses and from educational institutions. There are also possibilities for internal logical checks between different topics in the survey programme, such as literacy and educational qualifications.

147. In addition, special in-depth investigations could be undertaken subsequent to the main survey in order to test the validity of answers obtained in the original round of interviews. Such post-enumeration tests are of particular importance for evaluation of the quality of data on literacy. As was emphasized in section E above, the measurement of literacy in the usual type of survey and census operation is subject to important limitations. It cannot normally be expected that the time allowed for interviews and the qualifications of interviewers will be sufficient to determine the level of literacy of respondents with the desirable precision, especially when there is need for a clear distinction between simple literacy and functional literacy.

148. The most adequate method to check the reliability of responses concerning literacy obtained by means of routine interviews is to use standardized achievement tests for an appropriately chosen subsample of the survey population. With such tests conducted by specialized personnel and using uniform methods of administering and scoring the individual tests, the degree of literacy can be determined with satisfactory precision.

149. One example of an extensive standardized test of reading and writing ability for post-enumeration testing is from the 1953 population census in Yugoslavia (2.98). The major result of this in-depth investigation was that census data on literacy are not likely to be very reliable, but that the degree of precision may vary considerably in

connexion with specific conditions. Two main problems were identified. One concerned the extreme cases, complete illiteracy and very high literacy, which normally should not raise any problems but which might nevertheless be wrongly classified because of an inadequate system of enumeration. The other concerned intermediate cases which presented serious difficulties of a conceptual nature related to the lack of a criterion as to the level of ability to write and read at which literacy begins. In any case, an important conclusion of this post-enumeratin survey was that the chosen method of sampling control of literacy data did not raise any serious difficulties.

150. Another example of such a method for evaluation of literacy data relates to a post-enumeration study in connexion with a special survey of personnel in the Iranian Oil Company 1955, cited in (2.89). This study was based on reading and writing tests of a simpler type, but it is of particular interest because it was concerned with literacy in both the national language, Farsi, and in English. In both cases it was concluded that there was a fairly significant correlation between the census determination of literacy and test performance. It was, however, also found that comparability between literacy levels in the two languages was difficult to achieve, even with the use of tests. It is interesting to note that both studies concluded (contrary perhaps to general belief) that there was not a systematic bias towards overstating one's level of literacy.